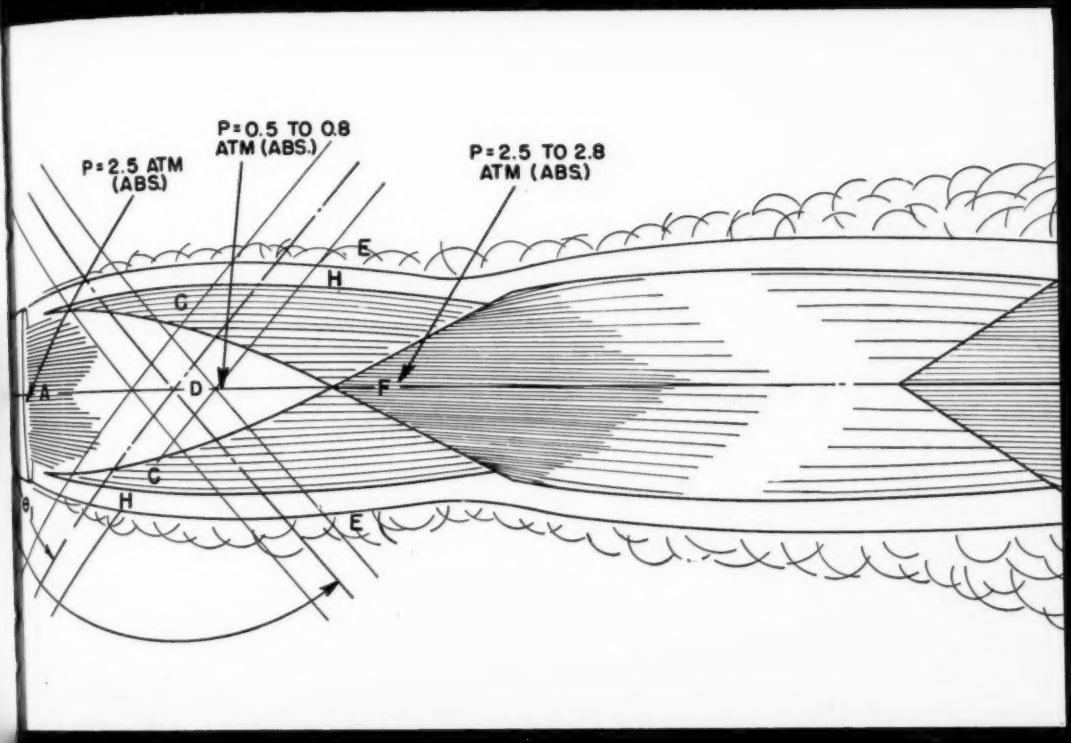
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 8 AUGUST, 1951

In This Issue

		Page
Microwave Filters Employing a Single Cavity Excited in More than One Mode	Wei-guan Lin	989
Conduction of Current in a Metallic Pipe Filled with a Conducting Liquid	H. Poritsky and H. Weil	1002
Gas Temperatures and Elastic Losses in Low Pressure Mercury-Argon Discharges	C. Kenty, M. A. Easley, and B. T. Barnes	1006
Pulse-Annealing for the Study of Relaxation Processes in Solids	W. E. Parkins, G. J. Dienes, and F. W. Brown	1012
Kinetics of Ordering in the Alloy AuCu	G. J. Dienes	1020
Flow of Gases in Porous Media	L. H. Wilson, W. L. Sibbitt, and M. Jakob	1027
Studies in Newtonian Flow. I. The Dependence of the Viscosity of Liquids on Temperature	Arthur K. Doolittle	1031
On Creep and Relaxation. III	B. Gross and H. Pelzer	1035
Electrostatically Induced Permanent Memory	Charles F. Pulvari	1039
Elastic Energy of an Ellipsoidal Inclusion in an Infinite Solid	Kenneth Robinson	1045
One-Sided Green's Function	Kenneth S. Miller	1054
Impedance of an Antenna over a Large Circular Screen	James E. Storer	1058
Expected Number of Crossings of Axis by Linearly Increasing Function Plus Noise	Frank W. Lehan	1067
Measurement of Velocity and Pressure of Gases in Rocket Flames by Spectroscopic Methods	F. P. Bundy, H. M. Strong, and A. B. Gregg	1069
Air Condensation in Hypersonic Flow	P. Wegener, S. Reed, Jr., E. Stollenwerk, and G. Lundquist	1077
Letters to the Editor:		
On a Theorem of Brouwer	H. F. Mathis	1084
Internal Friction of Cold-Worked Single Crystals of Copper	Ryukiti Robert Hasiguti and Tadamasa Hirai	1084
Volume-Temperature Relationships for the Room Temperature Transition in Teflon	Fred A. Quinn, Jr., Donald E. Roberts, and R. N. Work	1085
Work-Hardening of Mild Steel by Explosive Attack	John S. Rinehart	1086
On Kerr Cell Construction	W. R. Smythe	1087
Study of Oxide-Coated Cathode by X-Ray Diffraction Method	Eiso Yamaka	1087
X-Ray Detection of Long-Range Order in Ni_2Mn	B. L. Averbach	1088
Remark on the Composition Dependence of the Elastic Parameters of Beta-Brass	John S. Rinehart	1089
On Linear Expansion Coefficient and Melting Point of Metals	G. Bonfiglioli and G. Montalenti	1089
Announcements:		
Nuclear Physics in Europe		1090
1951 Awards for Essays on Gravity		1090
Index to Advertisers		xxvi

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
Lyle D. Borst
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Paul Kirkpatrick
F. W. Loomis
J. R. Oppenheimer
George B. Pegram
I. I. Rabi
Duane Roller
Harold K. Schilling
Frederick Seitz
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager and Assistant Treasurer*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Senior Accountant*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 22, Number 8 August, 1951

Microwave Filters Employing a Single Cavity Excited in More than One Mode*

WEI-GUAN LIN
Division of Electrical Engineering, University of California, Berkeley, California
(Received October 27, 1950)

A cavity resonator with input and output couplings represents a two-terminal-pair network with an infinite number of natural modes of oscillation. In some cavities of special shape, a number of degenerate modes with identical natural frequencies can be found. In a single cavity, various numbers of these degenerate modes can be coupled together to form a chain of coupled circuits by perturbing the otherwise ideal geometrical configuration of the cavity. The filter behavior is prescribed, and the two-terminal-pair network realizing this is obtained by a process of synthesis.

I. INTRODUCTION

IN a single cavity resonator there are an infinite number of modes. Suggestions on the possibility of utilizing some of them to form a multimode coupled circuit by means of a single cavity have been made,[1] with the advantage of saving space, weight, and cost. An experimental model of a two-mode filter has been described in the literature.[1] This employs two perpendicularly polarized TE_{111} modes to give the equivalence of two tuned circuits of a two-stage coupled circuit. One cavity or one iris is thus saved, but the single cavity is more complex. The full advantages of the principle can be realized only when we can control more than three modes in a single cavity.

The purpose of this report is to carry through the theoretical analysis to investigate the number of modes that is useful for our purpose of deriving a multimode filter.

II. A SINGLE CAVITY EXCITED WITH MORE THAN ONE MODE

1. Number of Possible Useful Modes in a Single Cavity

The interaction between the free oscillations of two tuned circuits depends on the coupling coefficient and on the ratio of their complex resonant frequencies. The closer the two frequencies are to each other, the less coupling is needed to transfer energy between the two circuits. In the problem of coupling together the various resonant modes in a single cavity resonator, the coefficient of coupling is inherently low, because, in an ideal region of simple geometrical configurations enclosed by perfectly conducting walls, orthogonality exists among the infinite number of modes, so that no energy transfer is possible among the modes. To produce coupling among the various modes, small perturbations in the otherwise ideal cavity configuration are introduced, and this perturbation must be small if small-perturbation theory is to be applied. Therefore, the modes of a single cavity which could be useful for our purpose of filter circuits must possess the same resonant frequencies or, in other words, should be degenerate modes.

For a rectangular cavity resonator bounded by perfectly conducting walls at $x=0, a$; $y=0, b$; $z=0, c$; the resonant wavelength for both the TM and the TE waves is given by

$$\lambda=[(l/2a)^2+(m/2b)^2+(n/2c)^2]^{-\frac{1}{2}}, \qquad (1)$$

where l, m, and n are the integer indices, one but not two of which may vanish. Moreover, $n=0$ is not allowed for the TE waves.

It can readily be seen that $a=b=c$ gives numerous numbers of degenerate modes with identical natural wavelengths. Because when λ is fixed, given by $2a/N$, $N^2=l^2+m^2+n^2$, and as N is made larger and larger, there exist more and more possible combinations of

* This work was done at the University of California Antenna Laboratory under ONR Contract N7onr-29529.
[1] G. L. Ragan, *Microwave Transmission Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), M.I.T. Radiation Laboratory Series, Vol. 9, pp. 673–677.

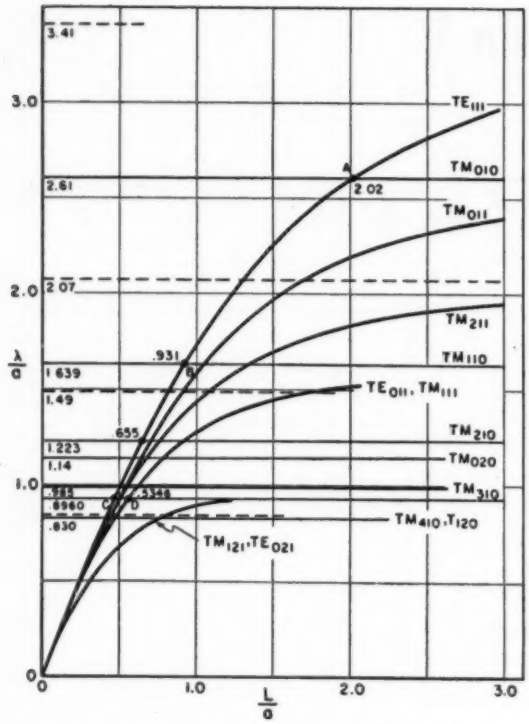


FIG. 1. Natural wavelengths of a circular cavity.

integers l, m, and n. For a given set of l, m, and n, there are six distinct permutations; this together with the fact that Eq. (1) characterizes both TE and TM waves makes the number of degenerative modes twelve when $n=0$, and ten when $n\neq 0$ (TE_{lm0} and TE_{ml0} are not allowed). However, we should observe that, for a given λ, the linear dimension a is larger the higher the order of degeneracy is; therefore, the number of modes that is practical is still limited.

For the next case of a cylindrical cavity bounded by the cylindrical surface $r=a$ and planes $z=0$ and $z=c$, the natural wavelength is given by

$$\frac{\lambda}{a}=\frac{2\pi}{[X_{lm}^2+n^2\pi^2/(c/a)^2]^{\frac{1}{2}}}, \tag{2}$$

where n is an integer, and X_{lm} is the mth root either of $J_l(X)$ for the TM waves, or of $J_l'(X)$ for the TE waves. Here $n=0$ is not allowed for the TE waves.

In Fig. 1, λ/a (of Eq. (2)) is plotted as a function of the ratio l/a for some of the lower modes, $n=1$. It is readily seen that any mode that is not axially symmetric is twofold degenerate, i.e., two identical modes with perpendicular polarizations. At A, $\lambda/a=2.02$; the two TE_{111}'s and the TM_{010} give another case of degeneracy. At points such as B and C, we have four modes of the same λ/a-ratios, i.e., two TE_{111} (even and odd) and two TM_{110} (even and odd) at B, and two TE_{111} and two TM_{120} at C. At point D, we have five modes: one circularly symmetrical TE_{011} mode, two TM_{111} modes with perpendicular polarizations, and two TM_{120} modes also with perpendicular polarizations. As we go further down in the graph, we will find many points where there are five modes of the same λ/a-ratio (wherever the functions $J_{0m}(X)$ and $J_{1m}'(X)$ have the same zero); however, no point can be found with more than five modes, for $n=1$. When n is allowed to take on values greater than one, the number of degenerate modes can be increased. For instance, at the point where the λ/a curves of the TE_{0mn} and the $TE_{0m'n'}$ meet, together with their dual TM modes, we have six degenerate modes. At this point, however, the size of the cavity is considerably increased, as can be seen by letting $m=2$, $n=1$, $m'=1$, and $n'=2$. Thus, we shall confine our investigation to cases where $n=1$.

For a spherical cavity, from

$$\left(\frac{\pi}{2r}\right)^{\frac{1}{2}}J_{n+\frac{1}{2}}(kr)P_n^m(\cos\theta)\begin{matrix}\cos m\phi\\ \sin m\phi,\end{matrix}$$

the basic wave solution for such a cavity, and from the fundamental boundary condition, we obtain the characteristic values of both the magnetic and the electric modes, respectively, with the following equations:

$$\begin{aligned} j_n(kr)&=(\pi/2r)^{\frac{1}{2}}J_{n+\frac{1}{2}}(kr)=0,\\ [krj_n(kr)]'&=0. \end{aligned} \tag{3}$$

These two equations possess distinct roots. Therefore, all the degenerate modes belong to the same type of waves, either magnetic or electric. For each type of wave, subject to condition (3), there are the even and odd modes in the ϕ-variation (except $m=0$). In the θ-variation, there are $n+1$ degenerate modes. Thus, for a particular characteristic number k_a, there may be as many as $2n+1$ degenerate modes.

The circular cavity is actually a limiting case of an elliptical cavity, so that the latter cannot be employed to give more degenerate modes. All other cavities of various geometrical configuration, such as coaxial or biconical cavities, likewise do not give more modes. They are, moreover, difficult to make.

2. A Single Cavity Excited with Two Modes

Let the circular cavity supporting the lowest mode be end-on coupled to two rectangular guides, supporting only the dominant mode, through small irises in the centers of its two end plates. The two broad faces of these guides are, however, 90° to each other. Theoretical and experimental[2] investigations point out that there exist two resonant modes in the cavity with perpendicular polarization; i.e., the angular variation is $\sin\phi$ for one and $\cos\phi$ for the other, each being in alignment with the exciting mode in the exciting guide. No coupling, then, exists between these two modes, nor

[2] W. W. Balwanz, "The resonant frequency of a cavity-type filter as a function of the size of the coupling iris," NRL Report R-3399, Naval Research Lab, Washington, D. C. (1949).

any between the input and the output guides in the ideal case free from any irregularity in the cavity.

Coupling will exist between them, however, when the ideal case is disturbed. To find the coefficient of coupling, we start with free oscillations. Assuming that only two modes, a and b, are excited, the field in the cavity is given by

$$\begin{aligned} \bar{H} &= A\bar{H}_a + B\bar{H}_a, \\ \bar{E} &= C\bar{E}_a + D\bar{E}_b, \end{aligned} \tag{4}$$

where

$$a = \int_v \bar{E}\cdot\bar{E}_a dv, \quad C = \int_v \bar{E}_a\cdot\bar{E}dv, \quad \text{etc.},$$

and $\bar{H}_a$ and $\bar{E}_a$ are normalized functions which have been defined by Slater[3] as follows:

$$\begin{aligned} \int_v \bar{E}_a\cdot\bar{E}_b dv &= \begin{cases} 1, & a=b \\ 0, & a\neq b \end{cases}, \quad k_a\bar{E}_a = \nabla\times\bar{H}_a, \\ \int_v \bar{H}_a\cdot\bar{H}_b dv &= \begin{cases} 1, & a=b \\ 0, & a\neq b \end{cases}, \quad k_a\bar{H}_a = \nabla\times\bar{E}_a, \end{aligned} \tag{5}$$

where k_a is the characteristic constant for the particular function.

Now we perturb the cavity boundary by pushing a small portion of the boundary surface into the cavity (by inserting a small screw, for instance). In the small volume between the original surface and the perturbed surface, the final $\bar{E}$ and $\bar{H}$ will be zero. There exists then a surface current, equal to $\bar{n}\times\bar{H}$, corresponding to the discontinuity of $\bar{H}$. Putting Eqs. (4) into the result of the method of orthogonal functions, Eqs. (III, 42, 43) of Slater's paper,[3] we have, over the small perturbation surface, for a time variation of $e^{j\omega t}$,

$$\begin{aligned} (\omega_a^2-\omega^2)A &= [\omega_a/(\mu_0\epsilon_0)^{\frac{1}{2}}]\int_{S'} \bar{n}\times\bar{H}\cdot\bar{E}_a da, \\ (\omega_b^2-\omega^2)B &= [\omega_b/(\mu_0\epsilon_0)^{\frac{1}{2}}]\int_{s'} \bar{n}\times\bar{H}\cdot\bar{E}_b da. \end{aligned} \tag{6}$$

If the perturbation is small, the perturbed field is very nearly equal to the original field, and we can then substitute the value of $\bar{H}$ from (4). We then write:

$$\int_{S'} \bar{n}\times\bar{H}\cdot\bar{E}_a da = A\int_{S'} \bar{n}\times\bar{H}_a\cdot\bar{E}_a da + B\int_{S'} \bar{n}\times\bar{H}\cdot\bar{E}_a da. \tag{7}$$

The integration is carried over the small perturbation surface S'. The integrations on the right can be transformed further by converting surface integrals into volume integrals and by expanding the divergence of the cross product of two vectors,

$$\begin{aligned} [\omega^2-\omega_a^2(1+m)]A - [\omega^2 p - \omega_a\omega_b q]B = 0, \\ -[\omega_b^2 p - \omega_a\omega_b q]A + [\omega^2-\omega_b^2(1+n)]B = 0, \end{aligned} \tag{8}$$

where we have written

$$m = \int_{V'} (H_a^2 - E_a^2)dv, \qquad n = \int_{V'} (H_b^2 - E_b^2)dv,$$

$$p = \int_{V'} \bar{H}_a\cdot\bar{H}_b dv, \qquad q = \int_{V'} \bar{E}_a\cdot\bar{E}_b dv.$$

When the two modes a and b are identical, $\omega_a=\omega_b=\omega_0$, and if the coupling screw possesses such symmetries that $m=n$, we have, from (8),

$$\begin{aligned} \omega^2 &= \omega_0^2[1+m\pm(p-q)] \\ &= \omega_0^2(1+m)\{1\pm[(p-q)/(1+m)]\}, \\ \omega &\cong \omega_0(1+m)^{\frac{1}{2}}\{1\pm[(p-q)/2(1+m)]\}, \end{aligned} \tag{9}$$

when p, q, $m \ll 1$.

Comparing this result with the conventional coupled circuits tuned to the same resonant frequency, it is seen that the modes in the cavity have a coefficient of coupling

$$k = (p-q)/(1+m). \tag{10}$$

The perturbation of the surface changes also the resonant frequency of each mode slightly.

Thus, we see that the input or the output guide together with its excited mode is represented by one-half of the circuit of Fig. 2(a). They would be independent of each other, were there no perturbation in the cavity proper. The coefficient of coupling M exists because of the insertation of a coupling screw, producing the required perturbation. This is then the equivalent circuit of our filter, b_1 and b_3 being the discontinuity susceptances between the input and output guides to the cavity, respectively. Figure 2(b) is ob-

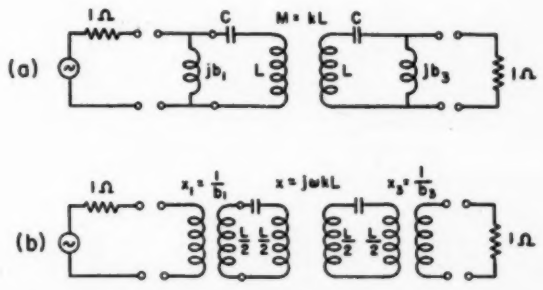


FIG. 2. The equivalent circuit of a single cavity excited with 2 modes.

[3] J. C. Slater, Revs. Modern Phys. Vol. 18, 441 (1946).

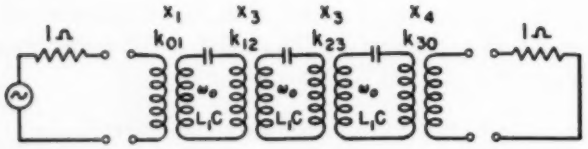


FIG. 3. The equivalent circuit of a single cavity excited with 3 modes.

tained from Fig. 2(a) by transforming the shunt coil into a loss-free transformer and by neglecting quantities of the order of $1/b_1$ in the series loops.

3. A Single Cavity Excited with Three Modes

At point A of Fig. 1 there are two TE_{111} modes and one TM_{010} mode with identical $\lambda/a=2.02$. Now, if the circular cavity of the preceding section (Sec. II, 2, Eq. (2)) is so scaled that the ratio λ/a is just 2.02, and if two coupling screws are inserted at each end of the cylindrical guide wall, each in line with the maximum electrical intensity of the input and the output rectangular guides, then in the round cavity there will be two orthogonal TE_{111} modes, each corresponding to one of the input and output guides, and one TM_{010} mode. There is no coupling between the TE_{10} mode in the rectangular guide and the TM_{010} mode in the round guide, because at the central hole, the tangential magnetic field of the latter mode vanishes, and former mode has no E_z. Also, there is no direct coupling between the two orthogonal TE_{111} modes when the coupling screws are so located and are 90° apart. Any coupling between the input and output circuits is through one TE_{111} to TM_{010}, and TM_{010} to the other TE_{111}. We thus establish an over-all system of a chain of three resonant circuits, represented by Fig. 3, where

$$k_{12}=(p_{12}-q_{12})/(1+m_{11}),$$
$$k_{23}=(p_{23}-q_{23})/(1+m_{22}), \tag{11}$$

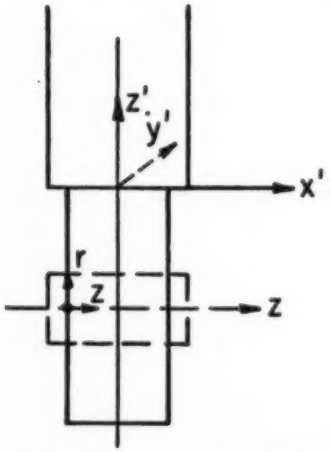


FIG. 4. Coupling between rectangular guides and the round cavity for a 4-mode filter.

TABLE I. Nonvanishing field components for the 4-mode cavity (point B of Fig. 1).

MODE / POINT	INPUT TE_{10}	(1) (e) TE_{111}	(2) (o) TM_{110}	(3) (e) TM_{110}	(4) (o) TE_{111}	OUTPUT TE_{10}
$(a,0,\frac{l}{2})$	$H_{x'}$	H_z	H_ϕ			
	$k_{01}\neq 0$					
$(a,90°,\frac{l}{2})$				H_ϕ	H_z	$H_{x'}$
					$k_{40}\neq 0$	
$(0,0,0)$		H_r	H_r	H_ϕ	H_ϕ	
		$k_{12}\neq 0$		$k_{34}\neq 0$		
$(0,90°,l)$		H_ϕ	H_ϕ	H_r	H_r	
		$k_{12}\neq 0$		$k_{14}\neq 0$		
$(a,45°,\frac{l}{2})$		H_z	H_ϕ	H_ϕ	H_z	
		$k_{14}\neq 0$		$k_{23}\neq 0$		

with

$$m_{ii}=\int_{\text{SCREWS}}(H_i^2-E_i^2)dv,$$
$$P_{ij}=\int_{\text{SCREWS}}\bar{H}_i\cdot\bar{H}_j dv, \tag{12}$$
$$q_{ij}=\int_{\text{SCREWS}}\bar{E}_i\cdot\bar{E}_j dv,$$

where $\bar{E}_i$ and $\bar{H}_i$ are the normalized field intensities evaluated at the tuning screw.

4. A Single Cavity Excited with Four and Five Modes

To correspond to B (also C) of Fig. 1, we take Fig. 4, where two rectangular guides are coupled to the cavity through two holes 90° apart on the cylindrical wall. The broad face of the former is parallel to the longitudinal axis of the latter. The two guides are perpendicular themselves. In Table I, we tabulate the nonvanishing field components at some favorable points

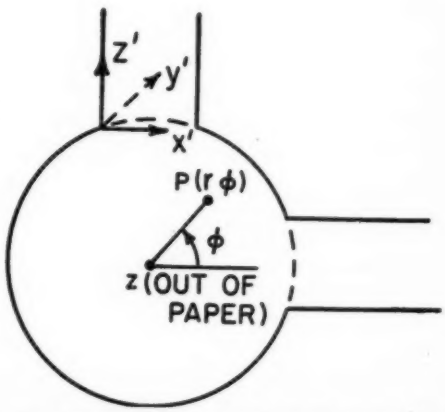


FIG. 5. Coupling between rectangular guides and the round cavity for a 5-mode filter.

where we are going to introduce the perturbations to effect the couplings for the modes of point B. The various coefficients of coupling are given by

$$k_{ij}=(p_{ij}-q_{ij})/(1+m_{ii}) \tag{13}$$

where m_{ii}, p_{ij}, and q_{ij} are given by (12).

The case of point D in Fig. 1 is realized by the arrangement shown in Fig. 5. The transverse magnetic field in the rectangular guide is aligned with the azimuthal component of the magnetic field of the TM_{120} mode in the round cavity. In Table II, the nonvanishing field components of the various modes at some preferable points are again given, and the coupling coefficients among the modes are examined.

TABLE II. Nonvanishing field components for the 5-mode cavity (point D of Fig. 1). ($J_1(x_{11})=J_1(3.832)=0$; $J_1(x_{12})=J_1(7.0156)=0$.)

MODE / POINT	INPUT TE_{10}	(1) (e) TM_{120}	(2) (e) TM_{111}	(3) TE_{011}	(4) (o) TM_{111}	(5) (o) TM_{120}	OUTPUT TE_{10}
$(a,0^\circ,\frac{\ell}{2})$	H_x'	H_ϕ	E_r	H_z			
	$k_{10}\neq 0$						
$(a,90^\circ,\frac{\ell}{2})$				H_z	E_r	H_ϕ	H_x'
						$k_{50}\neq 0$	
$(0,0^\circ,0)$		H_ϕ	H_ϕ	H_z			
		$k_{12}\neq 0$					
$(0,90^\circ,\ell)$				H_z	H_ϕ	H_ϕ	
					$k_{45}\neq 0$		
$(\frac{x_{11}}{x_{12}}a,90^\circ,0)$		H_ϕ	H_r	H_r			
			$k_{23}\neq 0$				
$(\frac{x_{11}}{x_{12}}a,0^\circ,\ell)$				H_ϕ	H_r	H_r	
					$k_{45}\neq 0$		

From Tables I and II, the exact equivalent circuits of the 4-mode and the 5-mode cavity are derived, respectively, and they are represented by Fig. 6 and Fig. 7.

5. Prescription of the Filter Characteristics

The equivalent circuits of a single cavity excited with various numbers of modes are all band-pass filters whose characteristics can be obtained from a prototype low-pass with a cut-off frequency of one radian per second, according to the principle of frequency normalization. The prototype structure of Fig. 8 is transformed into a chain of coupled circuits such as the one of Fig. 7 when the frequency transformation,

$$x=(\omega_0/W)[(\omega/\omega_0)-(\omega_0/\omega)] \tag{14}$$

is made, where ω_0 is the resonant frequency and W the bandwidth. When we identify Fig. 7 with Fig. 8, we

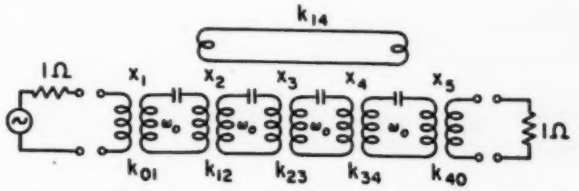


FIG. 6. The exact equivalent circuit of a 4-mode filter.

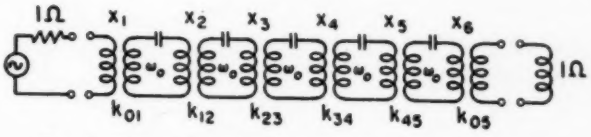


FIG. 7. The exact equivalent circuit of a 5-mode filter.

have

$$\begin{aligned} X_1^2&=(L/C_1)W,\\ X_{2k}^2&=LW^2/L_{2k}C_{2k-1}=X_1^4C_1^2/L_{2k}C_{2k-1},\\ X_{2k+1}^2&=LW^2/L_{2k}C_{2k+1}=x_1^4C_1^2/L_{2k}C_{2k+1},\\ X_{n+1}^2&=WLR/L_n=X_1^2C_1R/L_n \quad \text{for} \quad n \text{ even},\\ X_{n+1}^2&=WL/RC_n=X_1^2C_1/RC_n \quad \text{for} \quad n \text{ odd}. \end{aligned} \tag{15}$$

The insertion loss characteristic of Fig. 8 can now be prescribed; let

$$L_n=10\log(P_0/P_L)_n=10\log[1+h^2T_n^2(x)], \tag{16}$$

where $T_n(X)$ is the Tchebysheff polynomial of nth degree, and h^2 is an arbitrary constant which controls the tolerance in the pass band. When h is small, the maximum insertion loss in the pass band is

$$L_{\max}=10\log(1+h^2)=4.23h^2, \quad \text{for} \quad h^2\ll 1.$$

a. Two- and Three-Mode Filters

The actual values of the circuit elements of Fig. 8, for the cases of $n=2$ and $n=3$, can be solved by the elementary circuit theory. Thus (refer to Fig. 2(b)) symmetry consideration gives $X_1^2=X_3^2$, and, consequently,

$$\begin{aligned} R&=L_2/C_1,\\ h^2&=(1-R)^2/4R,\\ C_1^2&=2(1-R)/R. \end{aligned} \tag{17}$$

The choice of elements C_1 and R is controlled by the pass-band tolerance h^2.

For the case of $n=3$, symmetry requires $X_1^2=X_3^2$, or $R=C_1/C_3$ and $C_1=C_3$, or $R=1$. The expression for

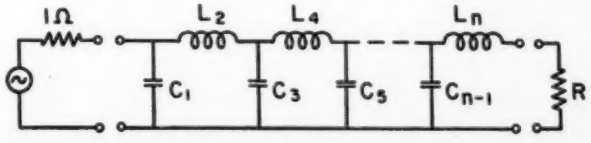


FIG. 8. A prototype low-pass filter.

P_0/P_L is

$$P_0/P_L=1+\tfrac{1}{4}[L_2C_1^2x^3-(2C_1-L_2)x]^2. \quad (18)$$

Identifying the coefficients of the quantity inside the bracket with those of $T_3(X)$, we have

$$L_2C_1^2=4,$$
$$2C_1-L_3=3,$$

one solution of which is (the other two roots form a complex pair)

$$C_1=C_3=2,$$
$$L_2=1. \quad (19)$$

To treat the cases of $n=4$ and 5, we define the power insertion loss P_0/P_L as

$$L=P_0/P_L=|\rho(j\omega)|^2, \quad (20)$$

where ρ is the voltage insertion loss ratio, i.e., the ratio between the voltages across the load resistance with the network interposed and with an ideal transformer inserted and adjusted for a match maximum power transfer; and the reflection coefficient is

$$\Gamma(p)=[1-Z(p)]/[1+Z(p)], \quad p=\sigma+j\omega, \text{ the complex frequency,} \quad (21)$$

where $Z(p)$ is the input impedance of the terminated network, Γ being related to the power insertion loss ratio by the following equation:

$$L=\frac{P_0}{P_L}=\frac{1}{1-|\Gamma(j\omega)|^2},$$

or

$$|\Gamma(j\omega)|^2=\frac{1}{(1+1/P_0/P_L)}; \quad (22)$$

and from (21),

$$Z(p)=[1+\Gamma(p)]/[1-\Gamma(p)]. \quad (23)$$

Since $Z(p)$ is positive real, $|\Gamma(p)|\leqslant 1$ for $\mathrm{Re}(p)\geqslant 0$. Also, $\Gamma(p)$ is analytic for $\mathrm{Re}(p)\geqslant 0$, because $1+Z(p)$ cannot vanish on the imaginary axis if $Z(p)$ is real positive. The poles of $\Gamma(p)$ and zeros of $\rho(p)$ are, therefore, confined to the left-half p-plane; thus, for a given $L(\omega)$, from (22), the poles of $\Gamma(p)$ and consequently $L(p)$ itself can be found if $L(\omega)$ has no finite poles. Similarly, $\rho(p)$ can be found from (20). Then $Z(p)$ is obtained from (23). The network realizing this $Z(p)$, and thus realizing the prescribed characteristic, is to be obtained by a process of synthesis.

For the four- and five-mode filters, let

$$L_4=1+\tfrac{1}{4}T_4^2(x)=1+\tfrac{1}{4}(8x^4-8x^2+1),$$
$$L_5=1+\tfrac{1}{4}T_5^2(x)=1+\tfrac{1}{4}(16x^5-20x^3+5x)^2, \quad (24)$$

so that the pass-band tolerance is about 1 db. By locating the roots of $\rho(p)$ from (24) and those of $Z(p)$ from (22), we have

$$\rho_4(p)=4(p^4+0.96p^3+1.46p^2+0.75p+0.28),$$
$$\rho_5(p)=8(p^5+0.944p^4+1.696p^3+0.983p^2+0.585p+1.25), \quad (25)$$

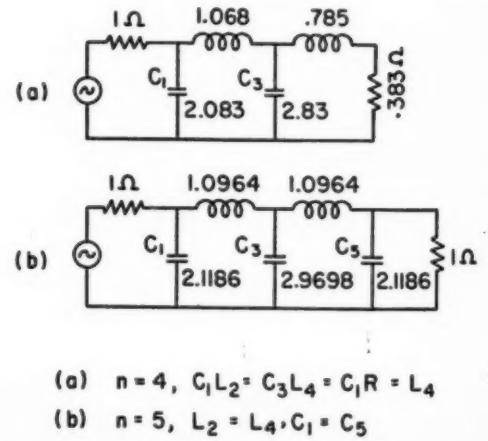
$$\Gamma_4=-\frac{p^4+p^2+0.125}{(p^4+1.46p^2+0.28)+(0.96p^3+0.75p)},$$
$$\Gamma_5=-\frac{p^5+1.25p^3+0.312p}{(0.948p^4+0.96p^2+0.125)+(p^5+1.701p^3+0.5897p)}. \quad (26)$$

To realize the networks giving these two reflection coefficients, we find that

$$1/Z_4=2.083p+1/1.068p+1/2.83p+1/2.83p+1/0.785p+0.383,$$
$$1/Z_5=2.1186p+1/1.0964p+1/2.969p+1/1.0964p+1/2.1186p+1, \quad (27)$$

respectively, for the cases of $n=4$ and $n=5$. The prototype structures in Fig. 9 correspond to Eqs. (27).

The dissipation is taken into account by assuming that the complex frequency p possesses a small but constant real part δ; and the dissipative insertion losses are as follows, neglecting terms involving δ of higher power than the third, $h=1$ for $n=2$ and $h=\frac{1}{2}$ for all



(a) $n=4$, $C_1L_2=C_3L_4=C_1R=L_4$

(b) $n=5$, $L_2=L_4, C_1=C_5$

FIG. 9. Prototype structures for the 4- and 5-mode filter.

the others,

$$L_2'=4\{(\delta^2+0.644\delta+0.707-x^2)^2+(0.644+2\delta)^2x^2\},$$
$$L_3'=4\{[\delta^3+\delta^2+1.25\delta+0.5-(3\delta+1)x^2]^2+x^2[3\delta^2+2\delta+1.25-x^2]^2\}, \quad (28)$$
$$L_4'=64\{[(5\delta+0.944)x^4-(10\delta^3+5.66\delta^2+5.09\delta+0.983)x^2+(1.696\delta^3+0.983\delta^2+0.585\delta+1.25)]^2+x^2[(4\delta+0.96)x^3-(2.88\delta^2+0.292\delta+0.75)]^2\},$$
$$L_5'=64\{[(5\delta+0.944)x^4-(10\delta^3+5.66\delta^2+5.09\delta+0.983)x^2+(1.096\delta^3+0.983\delta^2+0.585\delta+0.125)]^2+x^4[x^4-(10\delta^2+3.78\delta+1.69)x^2+(3.77\delta^3+5.09\delta^2+1.966\delta+0.585)]^2\},$$
$$\delta=R/L=(\omega_0/w)1/Q. \quad (29)$$

6. Coupling Elements of a Cavity Excited with Multimodes

Equations (14), (17), (19), and (27) give complete design data for our cavity filter. To obtain the actual coupling design, the value of L for each cavity must be evaluated explicitly. Let a cavity be formed by closing

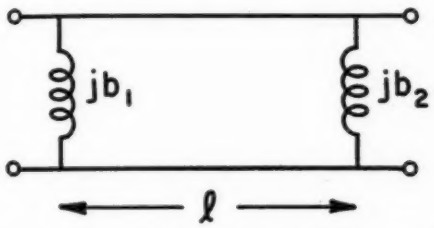


FIG. 10. The transmission line equivalent of a single cavity in wave guides.

a piece of wave guide by two end plates on which there are coupling devices, such as irises, etc. Then we have the circuit of Fig. 10 with open circuit admittances B_{11} and B_{22} as follows:

$$B_{11}=\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\left[1-\left(\frac{\omega_c}{\omega}\right)^2\right]^{\frac{1}{2}} \times\left\{b_1+\frac{b_2+\tan\left[\frac{l}{c}(\omega^2-\omega_c^2)^{\frac{1}{2}}\right]}{1-b_2\tan\left[\frac{l}{c}(\omega^2-\omega_c^2)^{\frac{1}{2}}\right]}\right\},$$
$$B_{22}=\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\left[1-\left(\frac{\omega_c}{\omega}\right)^2\right]^{\frac{1}{2}} \times\left\{b_2+\frac{b_1+\tan\left[\frac{l}{c}(\omega^2-\omega_c^2)^{\frac{1}{2}}\right]}{1-b_1\tan\left[\frac{l}{c}(\omega^2-\omega_c^2)^{\frac{1}{2}}\right]}\right\}. \quad (30)$$

At $\omega=\omega_0$, the resonant frequency, both B_{11} and B_{22} vanish. The proper length, l, of the cavity can be found from

$$\tan\left[\frac{l}{c}(\omega^2-\omega_c^2)^{\frac{1}{2}}\right]=\frac{b_1+b_2}{b_1b_2-1}\cong\frac{1}{b_1} \quad \text{for } b_2\gg b_1\gg 1, \quad (31)$$

and the slopes of B_{11} and B_{22} at $\omega=\omega_0$, their common zero, are given by

$$dB_{11}/d\omega|_{\omega=\omega_0}=n\pi[(b_1^2-1)/(\omega_0^2-\omega_c^2)^{\frac{1}{2}}](\epsilon/\mu)^{\frac{1}{2}},$$
$$dB_{22}/d\omega|_{\omega=\omega_0}=dB_{11}/d\omega|_{\omega=\omega_0}, \quad (32)$$

for large values of b_1 and b_2.

Now we know that, for a cavity coupled with two wave guides,[3,4]

$$L_1/\omega_0^2M_1=\tfrac{1}{2}dB_{11}/d\omega|_{\omega=\omega_0}$$

or

$$\frac{L_1}{X_1^2}=\tfrac{1}{2}n\pi\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\frac{b_1^2}{\omega_0(1-\omega_c^2/\omega_0^2)^{\frac{1}{2}}}. \quad (33)$$

Substituting the value for X_1^2 from (15), we have

$$b_1^2=(2C_1/n\pi)(\omega_0/w)(1-\omega_c^2/\omega^2)(\mu/\epsilon)^{\frac{1}{2}}(L_1/L)\times\{1-\omega_c^2/\omega_0^2\}^{-\frac{1}{2}}. \quad (34)$$

But L is already normalized, i.e.,

$$L=L_1/(\mu/\epsilon)^{\frac{1}{2}}[1-(\omega_c/\omega)^2]^{\frac{1}{2}}. \quad (35)$$

Therefore, the end result is

$$|b_1|=\left\{\frac{2C_1}{n\pi}\frac{\omega_0}{w}\left[1-\left(\frac{\omega_c}{\omega_0}\right)^2\right]\right\}^{\frac{1}{2}}=\frac{\lambda}{\lambda_g}\left\{\frac{2C_1}{n\pi}\frac{\omega_0}{w}\right\}^{\frac{1}{2}}. \quad (36)$$

We see that this value of b_1 is normalized with respect to the guide impedance that forms the cavity proper. However, for subsequent work we need the value of $|b_1|$ normalized with respect to the input and output guides, which for symmetrical operations are identical; therefore, we should write, in that case,

$$|b_1|=[\lambda_g^{(g)}\lambda/(\lambda_g^{(c)})^2]\{(2C_1/n\pi)(\omega_0/w)\}^{\frac{1}{2}}, \quad (37)$$

where $\lambda_g^{(g)}$ is the guide wavelength in the coupled guides, and $\lambda_g^{(c)}$ is that in the cavity proper. This change of scale is justified once we assume the pass band is small so that we can neglect the frequency sensitivity of the coupling susceptances.

From Eqs. (15) we find that

$$\begin{aligned}
&k=(W/\omega_0)[RC_1^2]^{-\frac{1}{2}} && \text{for 2-mode cavity}\\
&k_{12}=(W/\omega_0)[C_1L_2]^{-\frac{1}{2}} && \\
&k_{23}=(W/\omega_0)[L_2C_3]^{-\frac{1}{2}} && \text{for 3-mode cavity}\\
&k_{12}=k_{34}=(W/\omega_0)[L_2C_1]^{-\frac{1}{2}} && \\
&k_{23}=(W/\omega_0)[L_2C_3]^{-\frac{1}{2}} && \text{for 4-mode cavity}\\
&k_{12}=k_{45}=(W/\omega_0)[L_2C_1]^{-\frac{1}{2}} && \\
&k_{23}=k_{34}=(W/\omega_0)[L_2C_3]^{-\frac{1}{2}} && \text{for 5-mode cavity.}
\end{aligned} \quad (38)$$

[4] S. A. Schelkunoff, Proc. Inst. Radio Engrs. 32, 83 (1944).

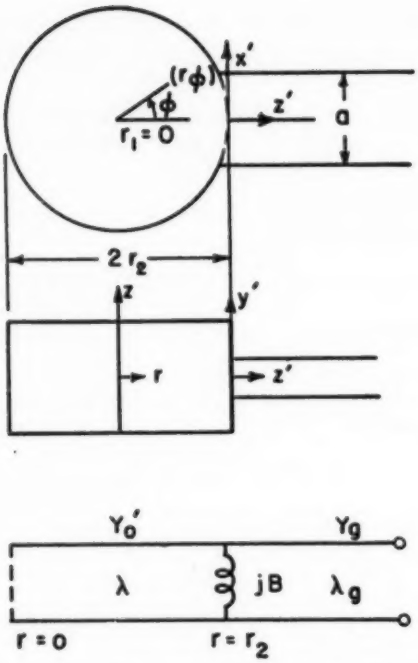


FIG. 11. Equivalent circuit of the coupling between a radial and a uniform transmission line.

7. Small Hole Coupling between Rectangular Guide and Circular Guide or Cavity

In our filters we use rectangular guides for the input and output lines, coupled to the circular cylindrical cavity by inductive irises either on the end-plates or on the cylindrical wall of the latter. Now we will give the discontinuity susceptance to be added to the transmission line circuit. We will assume that the coupling holes are small so that the field in the hole is constant and that the effect of the hole can be replaced by magnetic and electrical dipoles by following Bethe's line of reasoning in his studies[5] of small holes in an electromagnetic field.

a. End-On Coupling between Rectangular and Circular Guides

Let a wave be incident from the rectangular guide. In the circular guide there will be a transmitted wave through the hole, and a reflected wave is also set up in the rectangular guide. The nth normal mode H- and E-vectors in the guides are normalized:

$$\int_S |\bar{H}_{tn}|^2 dS = 1, \tag{39}$$

$$\int_S |\bar{E}_{tn}|^2 dS = Z_{0n}^2, \tag{40}$$

where $Z_{0n} = (\mu/\epsilon)^{\frac{1}{2}}(\lambda_g/\lambda)$ is the wave impedance for the mode which has been taken as a TE wave.

[5] H. A. Bethe, Phys. Rev. **66**, 66–163 (1944).

In the rectangular guide we have an incident wave and a reflected wave; in the circular guide we have a single transmitted wave. We then apply Lorentz's reciprocal theorem to the circular guide. In the hole the field is taken as the sum of the fields on both sides of the hole. The magnitude of the transmission coefficient is found,

$$T = 2[1 - (H_{at}^2/\bar{H}_{at}\cdot\bar{H}_{bt}) + j(-\lambda_g^{(b)}/2\pi\bar{H}_{at}\cdot\bar{H}_{bt})]^{-1}. \tag{41}$$

For small holes the discontinuity susceptance is large; thus, the normalized susceptance is given by

$$b = -(\lambda^{(b)}/2\pi M)(1/\bar{H}_{at}\cdot\bar{H}_{bt}) = 0.467[\lambda_g^{(b)}(\pi R^2 ab)^{\frac{1}{2}}/d^3], \tag{42}$$

where $M = 4/3(\frac{1}{2}d)^3$, and d = diameter of the hole. The field components H_{at}, etc., are defined by (40); quantities with subscript a refer to those in the transmitting rectangular guide, those with subscript b, to the receiving circular guide. The former has a width a and height b, while the latter has a radius R.

b. Coupling between a TE_{10} Mode in a Rectangular Wave Guide and a TM_{120} Mode in a Circular Cavity through a Circular Hole in the Side Wall of the Circular Cavity

Let the narrow side of the rectangular guide be in the direction of the longitudinal axis of the cavity, as shown in Fig. 11. In the rectangular guide there are three field components, H_z', E_y', and H_x', which will excite the H_ϕ, E_z, and H_r, respectively, in the circular cavity. We will consider the cavity as a radial transmission line supporting a TM_{120} mode alone, coupled through an inductive admittance to another uniform line supporting the dominant TE_{10} mode. In Fig. 11 we show the coordinates of the coupled system and the equivalent circuit. Energy is considered to be originated in the radial line, propagating through the coupling element to the uniform line.

Now the radius of curvature of the cavity wall is large compared with the linear dimension of the coupling hole, because we are handling only small holes compared with the wavelength, so the field distribution inside the cavity in the neighborhood of the side wall is the same as that in a rectangular guide of width $a' = \pi r_2$ and height $b' = l$. So, for the first approximation, we can treat our present small hole as one that couples together two rectangular wave guides of different transverse dimensions, and the result given by (42) is valid here;

$$b = -(3\lambda_g/d^3)(abr_2 l/4\pi)^{\frac{1}{2}}. \tag{43}$$

III. APPLICATION TO FILTER DESIGN

Before taking up sample designs, we will review the assumptions we have made or implied and discuss their justifications. There are four main assumptions: (a) the

isolation of the coupling discontinuity or iris; (b) the perfectly circular cavity surface; (c) parallel end plates perpendicular to the longitudinal axis of the cavity; and finally (d) small perturbations, i.e., small coupling holes ($d/\lambda \ll 1$) and small screws.

1. Isolation of Discontinuities

In the first case of our problem in which a round wave guide of $\frac{1}{2}\lambda_g$ in length is closed at both ends, the field intensities at one end plate of the higher mode generated at the other are being attenuated by a factor $\exp[-\alpha\lambda_g/2]$ with

$$\tfrac{1}{2}\alpha\lambda_g = \pi[(\lambda_c/\lambda_c')^2(\lambda/\lambda_c)^2 - 1]^{\frac{1}{2}}/[1-(\lambda/\lambda_c)^2]^{\frac{1}{2}}, \quad (44)$$

where $\lambda_c' = (2\pi/x_{mn})a$ is the cut-off wavelength of the high order mode.

For good isolation of the discontinuities at the two end plates, the value of (λ/λ_c), Eq. (44), should approach unity; i.e., the cavity is operated near the cut-off condition. However, under this operating condition, the Q-value of the cavity is very low. On the other hand, poor isolation of the irises on the end plates introduces the undesirable direct coupling and consequently decreases the attenuation in the stop-band. Hence, appropriate compromise must be made in choosing the operating condition.

In our experimental model of the 2-mode filter, we will so choose the diameter of the cavity that the Q-value is nearly maximum. The attenuation constant for the first higher order mode, TE_{01}, is estimated to be 5, which is considered as satisfactory for good isolation of the two discontinuities.

In both cases of the 3- and 5-mode filters, the dimensions are fixed for the coexistence of the degenerate modes. For the 3-mode case, the value of (44) is slightly less than 5. For the 5-mode filter, if we follow the line of reasoning postulated in Eq. (7), we have then a case of two discontinuities spaced by a distance $\frac{1}{2}\pi a$ in a rectangular wave guide of height $b=l$. Therefore, the ratio of the distance between the two discontinuities to the height of the guide is about 3, giving a value of the attenuation constant, Eq. (44), considerably larger than 5.

2. Ellipticity of the Cavity Wall

When the circular cavity possesses a slight ellipticity, the even and the odd modes will both suffer changes in natural wavelength, but by different amount.[6,7] To consider the allowable amount of change of the natural wavelengths of the various modes, we recall that, for small detuning, the detuning causes the relative transmission to decrease, and the effective coefficient of coupling to increase to the value given by

$$k_{eff} = [k'^2 + (\Delta/f_0)^2]^{\frac{1}{2}}, \quad (45)$$

where k' is the actual coefficient of coupling, and Δ is the difference between the resonant frequencies of the tuned circuits. The detuning will also make the attenuation curve of the tuned-circuit filter slightly unsymmetrical when the two tuned circuits do not have the same Q-values.

In order to realize the prescribed behavior of the filter, the effective value k_{eff} in Eq. (45) is set equal to the desired value of k for exact tuning. It is then necessary that

$$\Delta/f_0 < k \quad (46)$$

so that k_{eff} will not be greater than k. Equation (46), therefore, serves as the criterion of the amount of detuning and consequently the percentage of ellipticity allowable. Investigations into the actual modes employed bring out the fact that the ellipticity resulting from any reasonable manufacturing deviation in a cavity intended to be circular will not have serious effects on the behavior of the filter.

3. Parallel End Plates Perpendicular to the Cavity Axis

The deviation of the two end plates of the cavity from perpendicularity with the axis introduces again the frequency shift (and thus detuning among the modes) and the undesirable direct coupling among the modes.

Let the two end plates be given by

$$\begin{aligned} z &= d_1 + \alpha_1 x + \beta_1 y, \\ z &= d_2 + l + \alpha_2 x + \beta_2 y, \end{aligned} \quad (47)$$

where the α's, β's, and d's are small quantities measuring the degree of deviation from the ideal case. When they all vanish, we have the ideal case $z=0$ and $z=l$. To investigate the first effect, we make use of Eq. (9),

$$\Delta f/f = \tfrac{1}{2}m_{ii} = -\frac{1}{2}\int_{v'} (H_i^2 - E_i^2)\,dV, \quad (48)$$

$$dV = z\,dA = [d_1 + (\alpha_1 \cos\phi + \beta_1 \sin\phi)\Gamma]\,da,$$

where v' denotes the difference volume between the deformed cavity and the original unperturbed cavity; H_i and E_i are the normalized field components, evaluated at the unperturbed boundary enclosed by the difference volume:

$$\Delta f/f = (p^2\pi^2/k^2l^2)(d_1/l),$$

$$\begin{aligned} p &= 0, 1, 2, 3, \cdots \quad \text{for} \quad TM_{mnp} \\ p &= 1, 2, 3, \cdots \quad \text{for} \quad TE_{mnp}. \end{aligned} \quad (49)$$

Taking the second end plate into account, we have

$$\begin{aligned} \Delta f/f &= (p^2\pi^2/k^2l^2)[(d_1 - d_2)/l] \\ &= p^2(\lambda/2l)^2[(d_1 - d_2)/2]. \end{aligned} \quad (50)$$

[6] L. J. Chu, J. Appl. Phys. **9**, 583–591 (1938).
[7] L. Brillouin, Elec. Commun. **16**, 998 (1938).

Now if we take Eq. (2), for a small increment in the length l, we can derive the change in the natural wavelength and arrive at the same Eq. (50). Thus, the control of the tolerance in d_1 and d_2 is actually taken care of, by requiring the adherence to the tolerance of the length of the cavity. In our case of the 5-mode filter, by specifying that the relative detuning between the TM_{120} and the TM_{111} be less than the coefficient of coupling between these two modes, we find that the length of the cavity must be scaled to approximately the last thousandth of an inch.

Tilt adjustments, though not critical in causing frequency shift, nevertheless are important in the spurious coupling between certain modes, namely, between the TE_{0np} and the TM_{inp} modes. To make this point clear, we take Eq. (13) (the factor m_{ii} becoming zero for this case),

$$k_{ij}=\int_{v'}\bar{H}_i\cdot\bar{H}_j dv-\int_{v'}\bar{E}_i\cdot\bar{E}_j dv$$

$$=(\alpha/\sqrt{2}X_{11})(\lambda/l) \text{ for } TE_{011} \text{ and } TM_{111}^{(e)}$$

$$=(\beta/\sqrt{2}X_{11})(\lambda/l) \text{ for } TE_{011} \text{ and } TM_{111}^{(0)}. \quad (13a)$$

Let the actual tilts at the edge of the end plate at the X and the Y axis be, respectively, l_1 and l_2. Then $\alpha=l_1/a$ and $\beta=l_2/a$, and Eq. (13a) becomes

$$k=[1/\sqrt{2}(3.832)](\lambda/a)(l_1/l)$$

between TE_{011} and $TM_{111}^{(e)}$

$$=[1/\sqrt{2}(3.832)](\lambda/a)(l_2/l) \quad (51)$$

between TE_{011} and $TM_{111}^{(0)}$,

and the order of magnitude of tolerance in the tilt adjustments again can be estimated.

4. Sizes of Coupling Screws and Irises

In the theoretical analysis of the preceding sections, a simple theory of the unloaded cavity with no discontinuity was first established. Small perturbations were then introduced. The effect of the small perturbation was utilized to set up the filter circuit. One would immediately wonder as to how small the perturbation had to be in order for the field distribution to remain the same as that existing according to the simple theory. Experimental work shows that the sizes of coupling screws are not critical at all. That screws of reasonable size can be employed without any deteriorative effect can be justified from the fact that we deal mainly with sinusoidal and first-order and first-kind bessel functions, both of which are regular at their zeros and have fairly broad maxima. Further experimental evidence[2] indicates that an iris diameter as large as approximately three-tenths of wavelength (i.e., the narrow dimension of the usual rectangular wave guide) is suitable for transmission with negligible loss when using the TE_{011} and the TM_{111} modes. Therefore, Bethe's diffraction theory of small holes is believed to be useful at least up to this dimension.

5. The Design Problems

The design problems for the 3-mode filter will be discussed in detail to bring out the important point of frequency correction due to screws and irises, and the modification of the actual dimensions of the cavity to account for them.

First we take one of the two identical TE_{111}'s, in Eq. (31); let

$$l=p(\lambda_g/2)-\epsilon(\lambda_g/2\pi), \quad (52)$$

where ϵ is a small dimensionless number, then we have

$$\epsilon=-1/b_1. \quad (53)$$

When b_1 is negative or when the iris is inductive, l is less than p half-wavelengths. In our existing problem, the cavity is also coupled through a hole to a second rectangular guide whose broad face has been turned 90° with respect to that of the first guide. Therefore, the former guide is called upon to propagate the mode TE_{01} instead of the dominant TE_{10} mode; and therefore, as this second hole introduces only a small perturbation effect, we can then assume $b_2\gg b_1$. For the same reason, the two holes on the end plates and the two screws produce only perturbation effect to the TM_{010} mode. Thus, for both the TE_{111} and the TM_{010}, the corrected frequency is (there is only one perturbation hole to each of the two TE_{111} modes)

$$\omega_i=\omega_0[1+\tfrac{1}{2}(m_{ii}+C_{ii})] \quad i=1, 2\cdots, \quad (54)$$

with

$$m_{ii}=\int_{\text{screws}}(H_i^2-E_i^2)dv\cong\sum_{\text{screws}}(H_i^2-E_i^2)V_{si},$$

$$c_{ii}=2\sum_{\text{holes}}(MH_i^2-PE_i^2),$$

when the holes and the screws are small, and, consequently, the fields are nearly constant over the holes and the screws. Here $\bar{H}_i$ and $\bar{E}_i$ are the normalized fields to be evaluated at the holes and at the screws; M and P are given by Bethe as $4/3(\frac{1}{2}d)^3$ and $\frac{2}{3}(\frac{1}{2}d)^3$, respectively.

Thus, for either of the two TE_{111} and for the TM_{010}, we have, respectively,

$$m_{11}+C_{11}=(16.8/V)(\lambda/\lambda_g)^2\{(V_s/10)+(d^3/16)\},$$

$$m_{22}+C_{22}=(1/V)\{2V_s-0.573d^3\} \quad (55)$$

$$=2.5k-(0.0182/R|b|).$$

From these equations we find the relative detuning between the TE_{111} and TM_{010}, recalling that $\lambda=2.61a$, $\lambda_g=2l=4.04a$,

$$2\delta\omega/\omega=(m_{11}+C_{11})-(m_{22}+C_{22}), \quad (56)$$

$$\delta\omega/\omega=(1/2V)\{0.603d^3-1.296V_s\}, \quad (57)$$

$$\delta\omega/\omega=-0.778k+(0.0096/|b|R),$$

where R is in inches. To correct for this small detuning, we can lengthen the cavity slightly to decrease the resonant frequency of the TE_{111} mode. The change in length does not change that of the TM_{010} mode. If we take $l=\frac{1}{2}\lambda_g$ originally, then we have

$$dl/l=-(\lambda_g/\lambda)^2(d\omega/\omega)=-2.4(\delta\omega/\omega); \tag{58}$$

and the final length of the cavity is

$$l=\tfrac{1}{2}\lambda_g[1+(1/\pi b')+2.4(\delta\omega/\omega)], \tag{59}$$

where

$$b'=(\lambda_g^{(c)}/\lambda_g^{(g)})b,$$

$$\delta\omega/\omega=-0.778k+(0.0096/|b|R).$$

For the case of the 5-mode filter, (43) gives the discontinuity susceptance between the TE_{01} in the input or output line and the TM_{120} mode in the cavity.

We next will find the change of the natural wavelength of the mode TM_{120}. The condition for resonance is: the total admittance at $r=r_2+0$ (looking into the iris) should be zero. And we have

$$y+jb_1=0, \tag{60}$$

where y is the relative input admittance at $r=r_2$ of a radial transmission line open-circuited at $r=0$, and y is given by[8]

$$y=-j[J_1'(kr)/J_1(kr)] \tag{61}$$

or, substituting in Eq. (60), we have

$$\frac{J_1[2\pi r/\lambda]}{J_1'[2\pi r/\lambda]}=+\frac{1}{b_1}.$$

To solve for r/λ we introduce a small dimensionless number ϵ. We put

$$r_2=(7.02/2\pi)\lambda[1+(\epsilon/7.02)], \tag{62}$$

and we have

$$J_1(7.02+\epsilon)/J_1'(7.02+\epsilon)=+1/b_1. \tag{63}$$

Taking the Taylor series expansion of the bessel functions of the first order, at the neighborhood of the second root 7.02, for small values of ϵ, we have

$$\epsilon=1/(b_1-1/7.02). \tag{64}$$

Having corrected the radius r_2 for the TM_{120} mode, we have to adjust the length in order to make the modes TM_{111} and TE_{011} in tune with the former mode. To do this, we take, for the TM_{111} and the TE_{011},

$$\lambda=2\pi[(3.832/a)^2+(\pi^2/l^2)]^{-\frac{1}{2}}, \tag{65}$$

and to make $\Delta\lambda=0$, we must have

$$\Delta l=(3.832/\pi)^2(l^3/a^3)\Delta a$$

or

$$\Delta l/l=(3.832/\pi)^2(l/a)^2(\Delta a/a). \tag{66}$$

[8] Montgomery, Dicke, and Purcell, *Principles of Microwave Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), M. I. T. Radiation Laboratory Series, Vol. 8.

However, for each mode there is still the frequency change due to the perturbation screws and one or two perturbation holes. Thus, for the TM_{120} there are four screws and one hole. For both the TM_{111} and TE_{011} there are four screws and two holes. The fractional frequency changes of the three modes TM_{120}, TM_{111}, and TE_{011} are therefore, respectively, as follows:

$$\begin{aligned}(\delta\omega/\omega)_1&=0.525k_{12}+0.0965k_{23},\\ (\delta\omega/\omega)_2&=0.128k_{12}+0.51k_{23}-(0.0326/|b|),\\ (\delta\omega/\omega)_3&=+0.24k_{23}+(0.0022/|b|).\end{aligned} \tag{67}$$

When examined in this form, the fractional frequency changes, Eqs. (67), are seen to tend to be positive for the TM_{120} and the TE_{011} modes, but negative for the TM_{111} mode. By shortening the length of the cavity, the resonant frequencies of the TM_{111} and the TE_{011} modes can be raised toward that of the TM_{120}. Thus, if we take

$$\Delta l/l=-(2l/\lambda)(\Delta\omega/\omega), \tag{68}$$

where $\Delta\omega/\omega$ is taken as

$$\Delta\omega/\omega=(\delta\omega/\omega)_1-\tfrac{1}{2}[(\delta\omega/\omega)_2+(\delta\omega/\omega)_3], \tag{69}$$

the three modes are brought closer together.

The final dimension of this 5-mode filter is given as follows:

$$\begin{aligned}D&=(7.016/\pi)[\lambda+(\epsilon/7.02)],\\ l&=0.2673D[1+0.0606\epsilon-(2l/\lambda)^2(\Delta\omega/\omega)].\end{aligned} \tag{70}$$

The only problem which remains to be solved now is to find the value of b_1 from the lumped circuit equivalent prototype structure. Now looking into the cavity from the rectangular guide, the actual value of the open-circuited admittance is

$$B_{11}=Y_e(kr)\{b_1-[J_1'(kr)/J_1(kr)]\}, \quad k=2\pi/\lambda. \tag{71}$$

The slope of B_{11} at its zero, after being simplified by putting in (61), is

$$dB_{11}/d\omega|_{\omega_0}=Y_e(ka)(7.02b^2/\omega_0). \tag{72}$$

If now we take the first equation of (33), and then substitute the value of X_1 given by (15), remembering the normalization condition (35), we have

$$b_1=\{(\omega_0/w)(2C_1/7.016)\}^{\frac{1}{2}}\{Y_{10}^{TE}/Y_e\}^{\frac{1}{2}}, \tag{73}$$

where ω_0/w is ratio of the mid-frequency to the band width, C_1 is given in the prototype structure, and 7.02 is the second root of $J_1(x)$.

6. Experimental Models

Filter No. 1. A 2-mode filter, a single coupling screw, No. 4/40, $\omega_0/w=400$, $\omega_0=9375$ mc, $\lambda=3.20$ cm$=1.26$ in., $D=\frac{15}{16}$ in., pass-band tolerance 3 db, $R=0.172$, $C_1=3.140$ (Eq. (17)) $|b_1|=14.95$ (Eq. (36)), d_1 (diameter of the input and output iris)$=0.319$ in. (Eq.

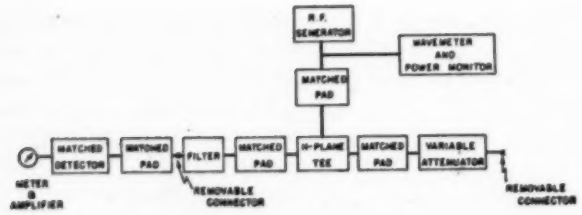


FIG. 12. Schematic experimental setup.

(42)), $l=1.005$ in. (Eq. (52)), $k=0.914\ V_s/V$ (V_s, volume of the coupling screw, Eq. (10)), $k=1.92\times10^{-3}$ (Eq. (38)).

Filter No. 2. A 3-mode filter, $\lambda=3.20$ cm, $\omega_0/w=400$, $\omega_0=9375$ mc, pass-band tolerance 1 db, $D=0.965$ in. (point A, Fig. 1), $L_2=1$, $C_1=2$ (Eq. (19)), $k=1.77\times10^{-3}$, $|b_1|=13.3$, $d_1=0.327$ in., $l=0.995$ in. (Eq. (59)).

Filter No. 3. A 5-mode filter, $\lambda=3.20$ cm, $\omega_0/w=454$, $w=20.7$ mc, $D=2.790$ in., $l=0.755$ in. (Eq. (70)), $|b_1|=16.56$ (Eq. (73)), $C_1=C_5=2.1186$, $C_3=2.9698$, $L_2=1.0964$ (Fig. 9), $k_{12}=k_{45}=1.445\times10^{-3}$, $k_{23}=k_{34}=1.220\times10^{-3}$ (Eq. (38)), $k_{12}=5.85V_1/V$, $k_{23}=1.132V_2/V$ (Eqs. (11), (12)), $d=0.382$ in. (Eq. (43)).

7. Measurements

In the experimental setup of Fig. 12, and H-plane tee is used to divide the power from the square wave modulated klystron oscillator equally between the two output branches, of which one contains the filter under test and the other a variable attenuator. The matched pads on the two sides of the tee are identical for balanced operation of the latter. In the measurement, at any one frequency the reading on the indication meter on the filter side is read. The detector together with the matched pad preceding and the indication meter following it are moved to the variable attenuator side; when adjusted to give the same reading on the indication meter, the reading of the variable attenuator gives the insertion loss of the filter.

The insertion loss curve of filter No. 1 is shown in Fig. 13. The band width at the maximum tolerance points is found to be about 21 mc, 14 percent less than the designed value of 23.4 mc, due to the presence of dissipation. At the low frequency end, over about 20 db, the attenuation is less than calculated, partly because of the direct coupling between the input and the output guides, but mainly due to the fact that the klystron signal generator is now operated far from its maximum power output point so that the signal is unable to override the noise in the circuit.

The insertion loss of filter No. 2 is shown in Fig. 14.

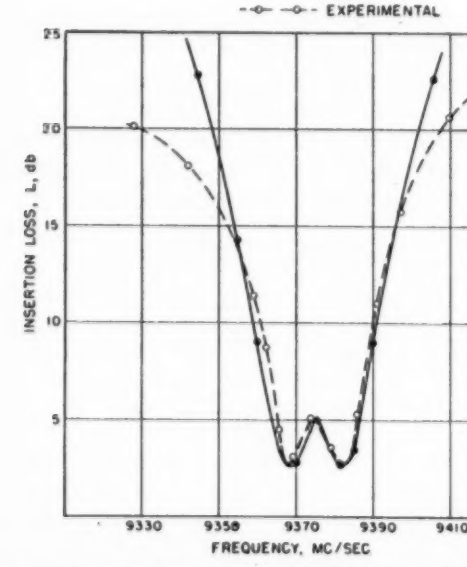


FIG. 13. Characteristics of a single cavity filter of 2 modes.

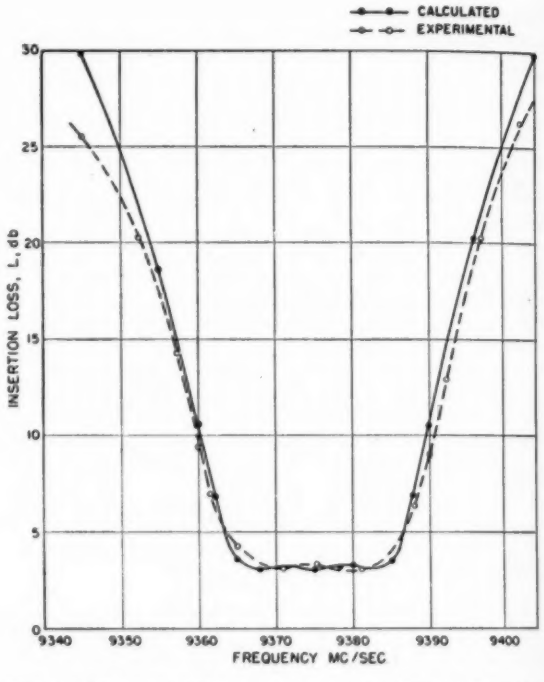


FIG. 14. Characteristic of a single cavity filter of 3 modes.

The insertion loss characteristic of filter No. 3 (see Fig. 16) is shown in Fig. 15; it is obtained by adjusting the four screws at the end plates equally in and out; i.e., the filter is adjusted for symmetrical operation.

Equations (28) are employed to check how well the experimental curves conform with the theory. In each experimental curve, the attenuation at the mid-frequency is read, this value being taken as L_n; and x, from Eq. (14), is made zero, and δ is solved for. This value of δ is substituted back into (28), and L_n is then plotted as a function of frequency by means of the frequency transformation (14). The agreement between the theory and the measurements is considered satisfactory for the crude experimental models.

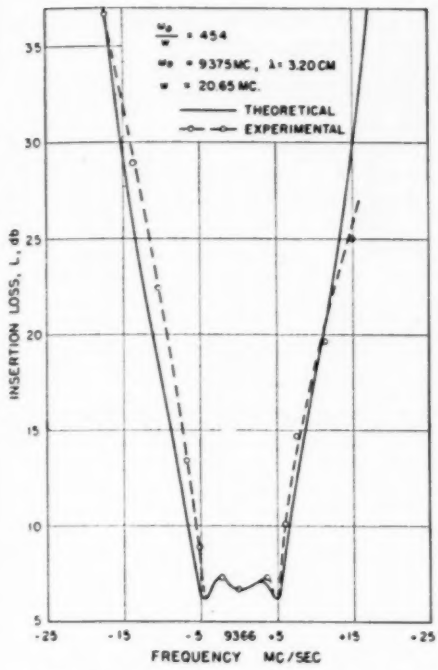


FIG. 15. Characteristic of a single cavity filter of 5 modes.

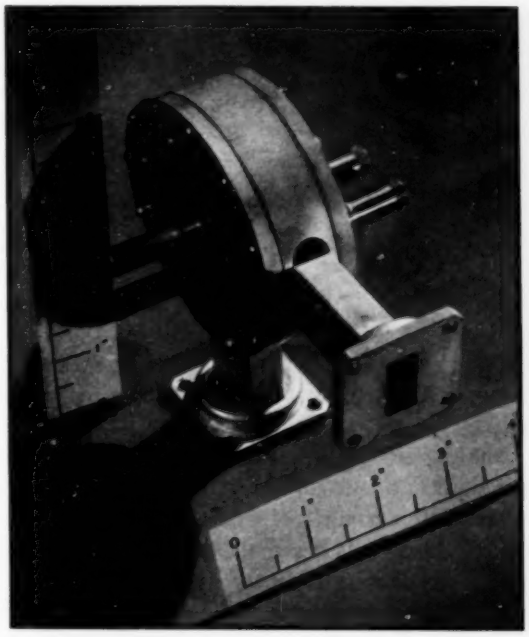


FIG. 16. Experimental model of the 5-mode filter.

IV. CONCLUSIONS

The experimental study has verified the theoretical analysis and the design calculations. It has shown that design equations can be relied upon to calculate frequency values well within one percent. When reasonable mechanical tolerance can be maintained, the theoretical prediction can be very closely realized. The tolerance, however, is by no means unduly critical, nor is the initial adjustment of the filter even when the point-by-point measurement has to be followed by means of the bench setup described in the text. The filter is light weight, compact, and easy to make.

We have repeatedly employed the concept of coupling coefficients in establishing the interrelations among the modes. To realize this coupling effect, small perturbations of the otherwise ideal cavity configuration are introduced. Conversely, if couplings among certain modes are to be surmounted, this analysis will enable us to know what kind of disturbances are to be avoided and how the coupling coefficient is to be raised or lowered.

The theory of coupling loops and irises between the feed line and a cavity resonator is far from being general and complete. Only under limiting assumptions can the coupling formulas for the iris be solved for some particular modes in the cavity and the dominant mode in the rectangular wave guide transmission line.

Various methods have been in existence for computing the discontinuity effect. Except for a few that arrive at the exact solution of the field problem, they all come to an estimate of the field over the plane of the iris. This estimation can be made by some mathematical approximations. It can just as well be made on a physical basis when mathematical approach is found difficult, as we have done in this report; it has been proved to be satisfactory.

ACKNOWLEDGMENT

Acknowledgment is due Dr. J. R. Whinnery and Dr. S. Silver, without whose encouragement and consultation this paper would not have been possible, and the staff of the Antenna Laboratory at the University of California for all the assistance they have rendered the writer.

Conduction of Current in a Metallic Pipe Filled with a Conducting Liquid

H. PORITSKY AND H. WEIL
General Engineering Laboratory, General Electric Company, Schenectady, New York
(Received November 22, 1950)

The electric potential across a metallic pipe is investigated when current enters and leaves through two diametrically opposite points and the pipe is either empty, filled with a conducting fluid, or lined with a thin film of the fluid.

I. INTRODUCTION

MEASUREMENTS of potential of the type investigated in this paper serve as a means of determining whether or not a vertical metallic pipe used for piping conducting fluids is or is not full. The possibility that, although the pipe has been emptied, a thin film of liquid adheres to the inner wall of the pipe is also taken into account.

The measurements are made as follows: An electric current I is allowed to enter and leave at two diametrically opposite points A and B. The potential difference is measured across two other points such as C and D in Fig. 1. By comparing the measured value of potential difference with the corresponding value computed by the formulas obtained in this paper, one can determine whether or not the pipe is full.

In Secs. II and V, the case of a completely full pipe is considered in detail. The completely empty pipe is a special case of the completely full pipe when the conductivity of the liquid vanishes. An additional approximate solution for the empty pipe is given in Sec. IV. The case in which a thin liquid film adheres to the pipe wall is considered in Sec. III and in the latter part of Sec. V. Results for the analogous two-dimensional potential problem are given in Sec. VI.

II. PIPE COMPLETELY FILLED WITH LIQUID

We begin by finding the potential V in the cylindrical pipe of outer radius b and inner radius a, but not taking into account current entering or leaving at the points A and B. By superposing product solutions of the laplace equation in the cylindrical coordinate system of Fig. 1, this potential is found to be

$$V_n = B_n(\alpha r)e^{in\theta}e^{i\alpha z}, \tag{1}$$

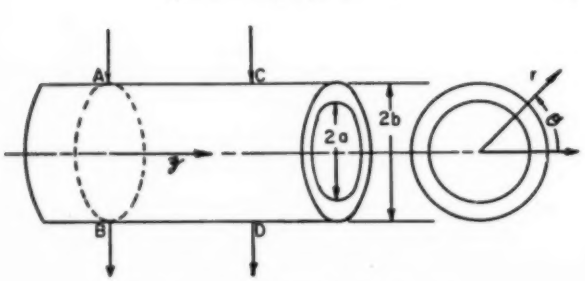


FIG. 1. Pipe and coordinate system.

where

$$B_n = I_n(\alpha r), \quad r<a,$$
$$B_n = I_n(\alpha r) - f\alpha a I_n(\alpha a) \times [I_n(\alpha a)K_n(\alpha r) - K_n(\alpha a)I_n(\alpha r)], \quad a<r<b. \tag{2}$$

In these equations

$$f = (\lambda_a/\lambda_b) - 1, \tag{3}$$

where λ_a and λ_b are the conductivities of the liquid and pipe, respectively, n is an integer, and α is an arbitrary real number. The constants in Eq. (2) were determined so that V_n satisfies the conditions that V and the radial current flow are continuous at $r=a$.

By superposition of solutions of the type (1), we next obtain a solution for the field which takes into account the current source and sink at A and B. We have

$$V = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} A_n(\alpha)V_n(\alpha r)d\alpha. \tag{4}$$

The coefficient function $A_n(\alpha)$ must be chosen so that Eq. (4) will correspond to the current flowing in at A and out at B. The current per unit area flowing into the cylinder normal to the boundary $r=b$, as determined by Eq. (5),

$$I' = \lambda_b(\partial V/\partial r)\,|_{r=b}, \tag{5}$$

is to be equated to the current input I' of the two point electrodes at A and B. Now, I' may be written in terms of the current I as

$$I' = I\delta(z)[\delta^*(\theta) - \delta^*(\theta-\pi)]/b, \tag{6}$$

where $\delta(z)$ is the unit impulse function in z, and $\delta^*(\theta)$ is the periodic impulse function in θ with impulses at $\theta=0, \pm 2\pi, \cdots$. The division by b is to obtain a current density per unit circumference rather than per radian. If $\delta(z)$ is resolved into a fourier integral and $\delta^*(\theta)$ into a fourier series, one can write I' as

$$I' = \frac{I}{2b\pi^2} \sum_{n=\pm 1, \pm 3, \cdots} \int_{-\infty}^{\infty} e^{ni\theta}e^{i\alpha z}d\alpha. \tag{7}$$

Equating Eq. (7) to Eq. (5) yields

$$A_n(\alpha) = I/2b\pi^2\lambda_b\alpha B_n'(\alpha b), \quad n \text{ odd},$$
$$A_n(\alpha) = 0, \quad n \text{ even}. \tag{8}$$

If we let

$$v = \alpha b, \quad z' = z/b, \quad \text{and} \quad \rho = a/b, \tag{9}$$

we thus have

$$V\Big|_{r=b} = \frac{I}{\pi^2 b\lambda_b} \sum_{n=1,3,\cdots} \cos n\theta \int_{-\infty}^{\infty} \frac{e^{ivz'} B_n(v)dv}{vB_n'(v)}, \quad (10)$$

where

$$B_n(v) = I_n(v) - f\rho v I_n'(\rho v) \times [I_n(\rho v)K_n(v) - K_n(\rho v)I_n(v)], \quad (11)$$

and

$$B'(v) = I_n'(v) - f\rho v I_n'(\rho v) \times [I_n(\rho v)K_n'(v) - K_n(\rho v)I_n'(v)]. \quad (12)$$

With the power series expansions for the bessel functions, one can check that, although $K_n(v)$ has a logarithmic singularity at $v=0$, the coefficient of $e^{ivz'}$ in the integrand of Eq. (10) has no singularity at $v=0$, and is, in fact, an analytic function of v^2 near $v=0$.

The integral is probably best evaluated by residues with the result

$$V\Big|_{r=b} = \frac{2I}{\pi b\lambda_b} \sum_{n=1,3,\cdots} \cos n\theta \sum_{k=1}^{\infty} \frac{e^{\pm v'_k z} B_n(v_k)}{v_k' D_n(v_k)}, \quad (13)$$

where $v_k = iv_k'$ are the roots of $B_n'(v)=0$, and

$$D_n = dB_n'(v)/dv.$$

These roots lie along the imaginary axis. Hence, the v_k' are real and are the roots of

$$2J_n'(v') + f\pi\rho v' J_n'(\rho v') \times [J_n(\rho v')Y_n'(v') - Y_n(\rho v')J_n'(v')] = 0. \quad (14)$$

The positive exponent is to be used in (13) for negative z and, conversely, the negative exponent for positive z.

The solution (10) or (13) evaluated for $\lambda_a=0$ gives the potential for an empty pipe.

III. PIPE WITH LIQUID FILM

If the pipe is not completely empty or completely filled with a conducting liquid, but instead has a thin liquid film adhering to the inner wall, a solution to the problem may be obtained by a procedure analogous to the one in Sec. II. Equations (10) and (13) for the voltage hold in this case at the outer edge of the pipe, but now we find

$$B_n(v) = C_n I_n(v) + D_n Y_n(v), \quad (15)$$

and

$$B_n'(v) = C_n I_n'(v) + D_n Y_n'(v),$$

where

$$C_n = -\rho v[I_n(\rho v)K_n'(\rho v) - I_n'(\rho v)K_n(\rho v)(f+1) - F_n K_n(\rho v)K_n'(\rho v)f], \quad (16)$$

and

$$D_n = -\rho v[I_n(\rho v)I_n'(\rho v)f - F_n(I_n'(\rho v)K_n(\rho v) - I_n(\rho v)K_n'(\rho v)(f+1))].$$

In these equations

$$F_n = -I_n'(\alpha c)/K_n'(\alpha c), \quad (17)$$

and c is the inner radius of the liquid film.

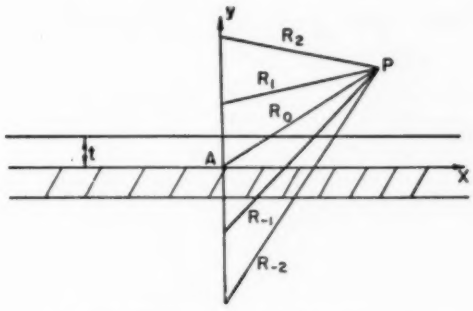


FIG. 2. Flat plate and string of images.

IV. AN APPROXIMATION VALID WHEN LIQUID CONDUCTIVITY VANISHES

If the conductivity of the liquid filling the pipe vanishes, and at the same time $(b-a)/b$ is small, the pipe may be thought of as straightened out into a flat plate, without introducing too much error in the potential field.

Once the pipe is replaced by a flat plate with the same thickness, t, as the pipe, we can use the method of images for calculating the resulting field. In the first place, the condition of no normal current flow at the outer pipe surface, which now becomes the upper plane boundary of the flat plate, may be satisfied by doubling the thickness of the flat plate, at the same time retaining the point source or point electrode through which the current enters at A, but allowing a double strength current $2I$ to enter there, half of which enters the lower half-plate, the other half the upper half-plate. The condition at the inner boundary of the pipe, which becomes the lower plane of the plate, can then be satisfied with the double thickness pipe material while maintaining the boundary condition at the outer surface of the pipe by placing positive images at distances $2nt$ above and below A.[1] This periodic array of point sources of current leads to the well-known series for the resulting potential

$$V_A(P) = \frac{I}{2\pi\lambda_b}\left[\cdots + \frac{1}{R_{-1}} + \frac{1}{R_0} + \frac{1}{R_1} + \frac{1}{R_2} + \cdots\right], \quad (18)$$

where the R_i are distances from the ith image point to any point, P, in space. The R_i are illustrated in Fig. 2 which shows a cross section of the flat plate perpendicular to the z direction of Fig. 1.

This series converges rather poorly, but the potential may be represented in the alternative form,[2]

$$V_A(P) = (I/2\pi t\lambda_b)[\text{const} + \ln(1/\rho) + 2\sum_{n=1,2,\cdots} K_0(n\pi\rho/t)\cos(n\pi y/t)], \quad (19)$$

in which $\rho = (x^2+z^2)^{\frac{1}{2}}$ is the radial distance from P to

[1] The method of images and the optical ray analogy, which will be used in Sec. V, are discussed in some detail in Sec. 4 of the paper by H. Poritsky, J. App. Mech. **16**, 123–133 (1949).

[2] E. Madelung, Physik. Z. **19**, 524 (1918).

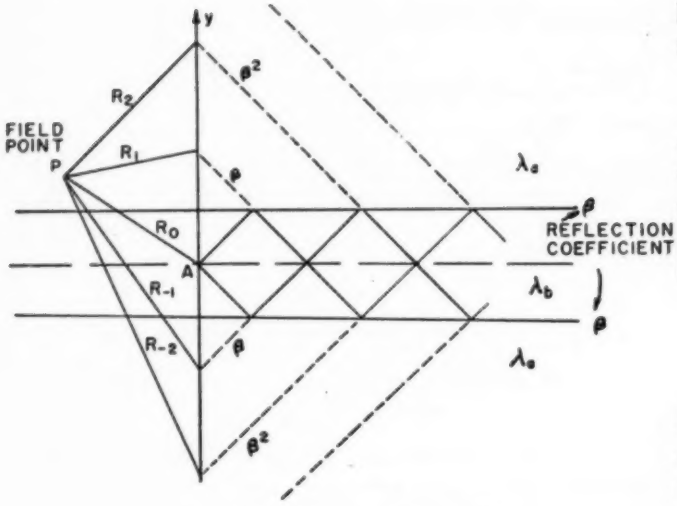


FIG. 3. Mirror analogy for full pipe.

the y axis of Fig. 2, and y is the distance in the y direction from P to the plane through A in Fig. 2.

To take into account the return electrode at B and also to take care of the periodicity of the original field in θ, one adds a series of fields of the form $\pm V_A$:

$$V=\cdots-V_{B_{-1}}+V_{A_0}-V_{B_0}+V_{A_1}-\cdots, \qquad (20)$$

corresponding to a displacement of each field by a multiple of half the average pipe circumference, $(1/2)\pi(a+b)$, in the x direction. At a distance ρ which is appreciable compared with $2t$, the series in Eq. (19) is negligible. Equation (20) reduces essentially to

$$V=\frac{I}{2\pi t\lambda_b}\sum_{n=-\infty}^{\infty}\left[\ln\frac{1}{\rho_{A_n}}-\ln\frac{1}{\rho_{B_n}}\right], \qquad (21)$$

where ρ_{A_n} and ρ_{B_n} are the distances from the y axes and

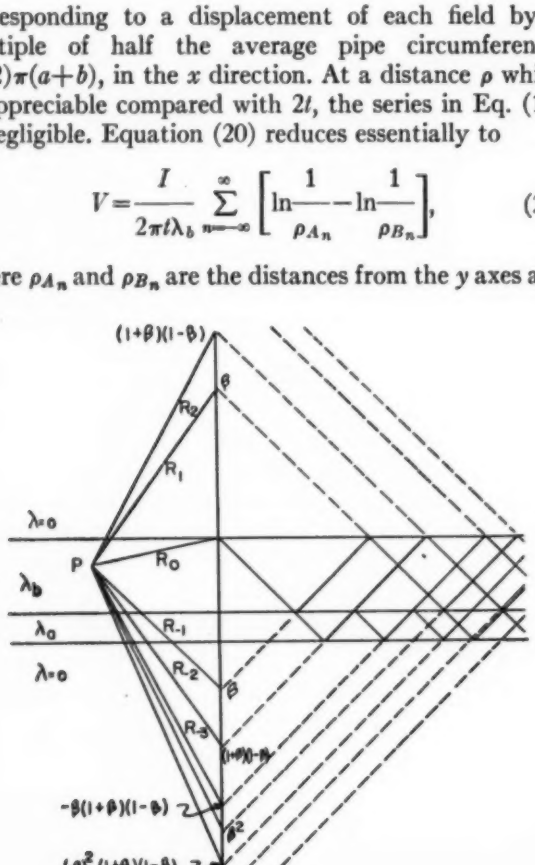


FIG. 4. Mirror analogy for pipe with a liquid film.

lines parallel to it, through the points A_n and B_n, respectively. This can be summed into

$$V=-(I/2\pi t\lambda_b)\,\mathrm{Re}\{\ln\tan[(x+iy)/(a+b)]\}, \qquad (22)$$

where "Re" is the "real part of." As stated, this is not valid near A; but there Eq. (22) takes care of the constant and log terms in Eq. (19) and of the field due to all the images, leaving the K_0 series in Eq. (19) to be added.

V. APPROXIMATION VALID NEAR THE POINT ELECTRODE

Near A, the convergence of the previous solution (13) is very poor. It is of interest, therefore, to develop some other, even if only approximate, solution which is valid near A.

Let us again neglect the curvature of the pipe near A, thus replacing the pipe by a flat plate and the cylindrical liquid by a liquid filling the semi-infinite space below it. The solution will not be valid for distances from A much larger than t.

Reflection across $y=0$ in Fig. 2 is still possible. This doubles the plate thickness and also fills the space above and below the plate with a liquid of conductivity λ_a.

In this case, the following series, similar to Eq. (18), is valid for the potential V within the plate:

$$\frac{2\pi\lambda_b V}{I}=\frac{1}{R_0}+\beta\left(\frac{1}{R_1}+\frac{1}{R_{-1}}\right)+\beta^2\left(\frac{1}{R_2}+\frac{1}{R_{-2}}\right)+\cdots, \qquad (23)$$

where

$$\beta=(\lambda_b-\lambda_a)/(\lambda_b+\lambda_a), \qquad (24)$$

and R_0, R_1, R_2, $\cdots$ are as in Fig. 2. This series reduces to $1/R_0$ for $\beta=0$, i.e., when $\lambda_a=\lambda_b$.

To obtain (23), we superpose axially symmetric product solutions of the potential equation so that

$$\frac{2\pi\lambda_b V}{I}=\frac{1}{R_0}+\int_0^\infty A(\alpha)\cosh\alpha y J_0(\alpha\rho)d\alpha,$$
$$\text{for}\quad |y|<t; \qquad (25)$$

$$\frac{2\pi\lambda_b V}{I}=\int_0^\infty B(\alpha)e^{-\alpha|y-t|}J_0(\alpha\rho)d\alpha,$$
$$\text{for}\quad |y|>t, \qquad (26)$$

where $R_0=(\rho^2+y^2)^{\frac{1}{2}}$. The $1/R_0$ term exhibits the singularity at the point source, while the integral terms will be used to satisfy the boundary conditions at $y=\pm t$. The forms of the integrands express the fact that the potential vanishes at infinity and that it is symmetric in y. Because of this symmetry, it suffices to consider boundary conditions at the upper boundary only. At the boundary, V and $\lambda(\partial V/\partial y)$ must be continuous.

To apply these conditions, it is convenient to express

$1/R_0$ in the form,

$$(1/R_0)=\int_0^\infty e^{-\alpha|y|}J_0(\alpha\rho)d\alpha \tag{27}$$

so that the boundary conditions simply require that

$$\begin{aligned} e^{-\alpha t}+A\cosh\alpha t&=B,\\ -e^{-\alpha t}+A\sinh\alpha t&=-(\lambda_a/\lambda_b)B. \end{aligned} \tag{28}$$

Now $A(\alpha)$, obtained by solving these equations, may be substituted in Eq. (25) for V, resulting in the expression

$$\frac{2\pi\lambda_b V}{I}=\frac{1}{R_0}-\int_0^\infty \frac{e^{\alpha y}+e^{-\alpha y}}{1-(1/\beta)e^{2\alpha t}}J_0(\alpha\rho)d\alpha. \tag{29}$$

When the denominator of the integrand in Eq. (29) is expanded in a series of inverse powers of $e^{2\alpha t}$, term by term integration leads to the expansion (23).

Again the return electrode should be taken into account, as in Eq. (20).

As in Fig. 2, the series (23) may be interpreted as the resultant field due to a source at A plus an infinite string of image sources along the y axis through A. The positions of these images may be obtained by means of the optical ray analogy shown in Fig. 3, where two rays of an arbitrary slope $\pm m$ are drawn through the point A, and continued till they hit the lines $y=\pm t$, where they are reflected, and so on. When the rays are continued backwards (as shown by the dotted lines), they intersect the y axis at the position of each image point source. The "strength" or "intensity" of each image is obtained by multiplying each ray by a "reflection coefficient" β upon reflection and by a "transmission coefficient" $(1+\beta)$ upon possible transmissions.

Figure 3 could be modified by retaining the surface $y=0$ and the original thickness of the wall, provided a reflection coefficient $\beta'=1$ be associated with $y=0$. The images are obtained in an analogous way by projecting back a ray leaving the source at A with an arbitrary negative slope.

Again in the case of the pipe with a liquid film, if the pipe is thin enough, it may locally be considered as straightened out and V obtained in a small region near A by the image method. In this case there are three reflection surfaces corresponding, respectively, to the air-liquid, liquid-air, and metal-air interfaces. With each of these is associated a reflection coefficient σ and a transmission coefficient $1+\sigma$. For the two air surfaces, $\sigma=1$. For the liquid-metal interface, however,

$$\sigma=\beta=(\lambda_b-\lambda_a)/(\lambda_b+\lambda_a) \tag{30}$$

for reflections of rays incident from above, and $\sigma=-\beta$ for reflections of rays incident from below.

Part of the ray system and resulting series of images is shown in Fig. 4, with each image labeled with the proper intensity. From the figure one sees that the net potential is given by

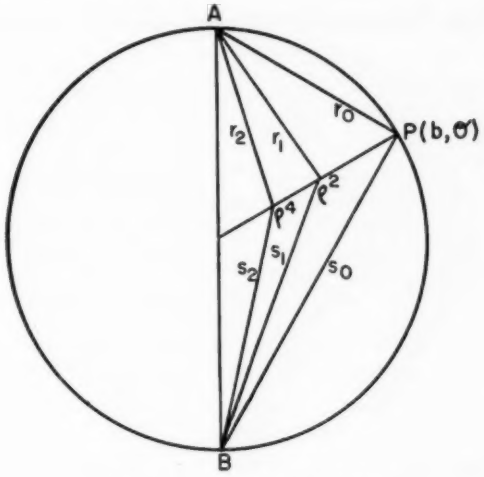


FIG. 5. Notation for Eq. (32).

$$\frac{2\pi\lambda_b}{I}V=\frac{1}{R_0}+\beta\left(\frac{1}{R_1}+\frac{1}{R_{-1}}\right)+(1-\beta^2)\left(\frac{1}{R_2}+\frac{1}{R_{-2}}\right)$$
$$-\beta(1-\beta^2)\left(\frac{1}{R_3}+\frac{1}{R_{-3}}\right)\pm\cdots. \tag{31}$$

VI. TWO-DIMENSIONAL FIELDS

It is much simpler to calculate the potentials of two-dimensional fields, where the current J per unit axial length enters along a line AA parallel to the axis, and leaves along a diametrically opposite line BB. Of course, the potential measurement must now be carried out for diametrically opposite lines corresponding to a θ different from zero.

The potential is found to be

$$V=\frac{2J}{\pi\lambda_b}\left[\ln\frac{r_0}{s_0}+2\beta\ln\frac{r_1}{s_1}+2\beta^2\ln\frac{r_2}{s_2}+\cdots\right], \tag{32}$$

where J is the input current density per unit length, and r_n and s_n are the distances from the points (ρ^{2n},θ) to A and B as shown in Fig. 5. From Sec. II it will be recalled that ρ is the ratio a/b of the inner pipe diameter to the outer one. This result may be obtained by properly combining product solutions in a manner somewhat analogous to the method in Sec. II. Alternatively, it may be found by mapping the pipe and liquid conformally on a plane sheet and semi-infinite space below it, by means of the conformal transformation from the x, y to a ξ, η plane,

$$\begin{aligned} \eta&=\ln[(x^2+y^2)^{\frac{1}{2}}/b],\\ \xi&=\theta, \end{aligned} \tag{33}$$

and then using images and fourier integrals in a method similar to the one in Sec. V.

Gas Temperatures and Elastic Losses in Low Pressure Mercury-Argon Discharges

C. KENTY, M. A. EASLEY, AND B. T. BARNES
General Electric Company, Lamp Development Laboratory, Nela Park, Cleveland, Ohio
(Received January 22, 1951)

The increase ΔT in average gas temperature for ac and dc has been determined from measurements of the increase in pressure. A small McLeod gauge was used. Wall temperature was regulated. Corrections for end effects were made by comparing results for long and short tubes of 3.6 cm diameter. Assuming the radial variation of heat input to be parabolic, the temperature distribution was calculated using published values of heat conductivity K. From the results the heat input P_m per centimeter length was calculated. This value was compared with the elastic loss P_e computed from electron temperature T_e and number of electrons N_e per centimeter column, taking into account the variation of mean free path with electron velocity. Good agreement was obtained for mercury pressures ranging from 1 to 25 microns, with the arc current 0.2–0.6 ampere and the argon filling pressure 3.5 mm (Hg).

THE heating (above wall temperature) of the gas in a discharge tube is, provided the gas is monatomic, almost entirely due to the energy losses of the electrons in elastic collisions with the atoms. Since these losses may represent a considerable fraction of the total wattage input, a knowledge of them is necessary in any study of the energy balance of the discharge.

If the electron temperature and concentration and the cross section for elastic collisions are known, these losses may be calculated. They may also be determined from the heat conductivity of the gas if the radial temperature distribution of the gas is known. This temperature distribution may be obtained from measurements of the increase in pressure incident to the operation of the discharge.

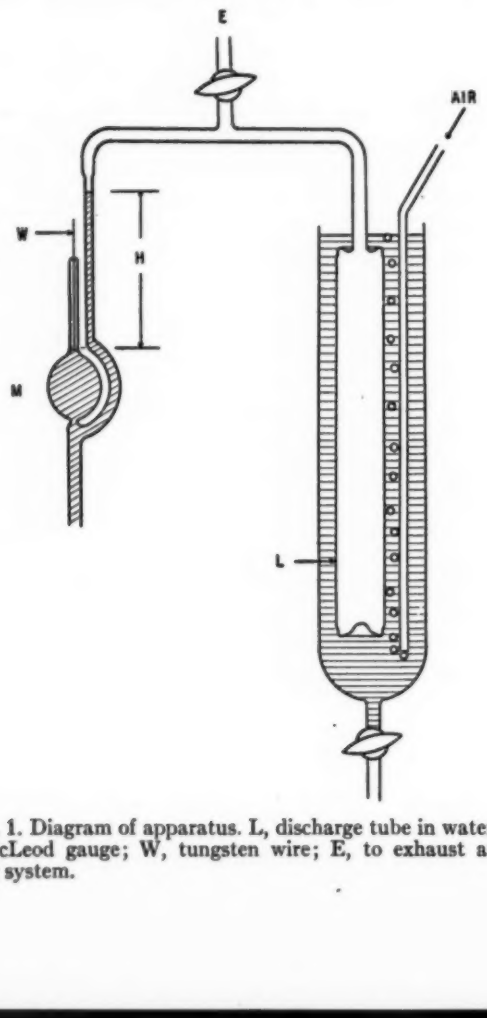


FIG. 1. Diagram of apparatus. L, discharge tube in water bath; M, McLeod gauge; W, tungsten wire; E, to exhaust and gas dosing system.

An attempt was made by Sommermeyer,[1] using both methods, to determine the elastic energy losses in the positive column of low pressure discharges in the rare gases argon, neon, and helium. Sommermeyer obtained only an approximate agreement between results using the two methods, the energy loss determined from the pressure increase exceeding the energy loss calculated from collisions of electrons with atoms by as much as 100 percent. However, his knowledge of the energy distribution and concentration of the electrons was incomplete, inasmuch as the discharges had running striations. There were also difficulties with gas impurities. A basic source of error was failure to take proper account of the variation of mean free path with electron velocity.

With Hg added to the rare gas, there is the great advantage of the absence of striations over a wide range, and relatively exact probe measurements of electron temperature and concentration are possible. In what follows, a study of the elastic losses in a mercury-argon discharge will be described.

APPARATUS AND METHOD

The apparatus is shown schematically in Fig. 1. Here the lamp L is a standard 40-watt fluorescent lamp, approximately 1½ inches in diameter and 4 feet long. It contains a drop of Hg and any desired pressure of argon or other rare gas. The lamp is surrounded by a water bath. With bubble stirring, departures from uniform temperature were less than ¼°C. For temperatures below 0°C, an acetone bath cooled by dry ice was used.

Connected to the lamp is a McLeod gauge[2] M. Besides having a small volume (8 cm³) and etched capillaries, this gauge has a special feature—a tungsten wire

[1] K. Sommermeyer, Ann. Physik **13**, 315 (1932); **28**, 240 (1937). These papers contain references to work done by others.

[2] K. Sommermeyer, reference 1, used a capacitance membrane manometer and later a Simon and Ruhemann Hg micrometer manometer for measuring pressures.

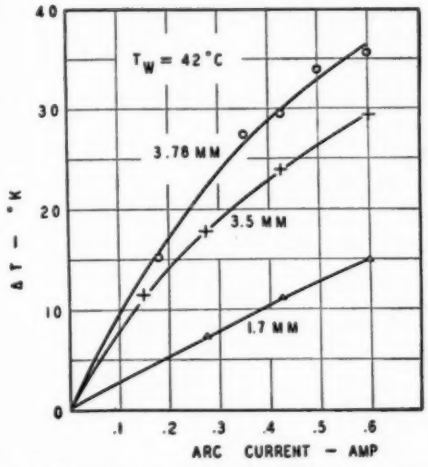


FIG. 2. Heating effect as a function of discharge current for three pressures of argon. ΔT is the increase in average gas temperature caused by the discharge. t_w is the wall temperature.

sealed through the top of the closed capillary and extending nearly to the bottom of it. When the Hg slowly rises and makes contact with the tip of this wire, it actuates a delicate relay which closes another relay and produces an audible click. At the time of this click, the height H of the Hg in the open capillary is read. Using the gauge in this way, and for its particular constants, the pressure is given by

$$p=0.0283H, \tag{1}$$

where p and H are both in millimeters. The gauge operated very satisfactorily in its designed range of 1–6 mm. In the middle of the range, the relative error was estimated to be $<\frac{1}{4}$ percent and the absolute error <1 percent. In the present experiments, practically only the relative error is involved. Further details of the gauge will be given elsewhere.

The ratio of the average gas temperatures (in °K), with and without the discharge, will be given approximately by

$$T_2/T_1=p_2/p_1. \tag{2}$$

It can be shown that "average temperature" as here used means the average taken over the groups of molecules and not over the volume.

Actually, Eq. (2) will not be quite true because of a small displacement of gas out of the discharge tube and into the gauge and its connecting tubing (of combined volume V_g) which takes place when the discharge is started. This can be taken care of by noting that the total quantity of gas is constant; i.e., that $\sum(pV)/T$ is constant. Thus, if T_g is the absolute temperature of the gauge and its connecting tubing, it follows that

$$p_1[(V_L/T_1)+(V_g/T_g)]=p_2[(V_L/T_2)+(V_g/T_g)], \tag{3}$$

where V_L is the volume of the lamp. On rearrangement, Eq. (3) becomes

$$T_1/T_2=(p_1/p_2)[1+(V_g/V_L)(T_1/T_g)] -(V_g/V_L)(T_1/T_g). \tag{4}$$

For the discharge tubes used,[3] $V_L=1200$ cm³ and $V_g=40$ cm³ ($V_g/V_L=0.033$), and the correction is negligible, i.e., <0.6 percent in T_1/T_2 for all cases and $<\frac{1}{4}$ percent for most conditions worked with. However, for a short lamp, to be considered below, the correction is considerable.

The procedure for measuring pressures was to first read p_1, then start the discharge, and after about 20 seconds read p_2. Tests showed this time to be ample for the attainment of the new equilibrium. Then the discharge was shut off, and after 20 seconds p_1 was read again. The two values of p_1 showed no significant difference and were averaged. Water temperatures were read at the time of all pressure measurements. These usually changed less than $\frac{1}{2}$°C, but the average temperature was used (for T_1) to correspond to the average p_1.

RESULTS

To render the results clearer and more useful, the rise in average gas temperature, caused by the discharge, has been expressed as $\Delta T=T_2-T_1$. In Fig. 2, values of ΔT have been plotted against rms current for three different filling pressures of argon and for a wall temperature t_w of 42°C. Filling pressures are at 27°C. Unless otherwise stated, the discharge was operated on ac with inductive ballast. Auxiliary cathode heating was employed only at the time of starting. It will be seen that ΔT rises somewhat less rapidly than the current.

The variation of ΔT with wall temperature is shown in Fig. 3. It will be seen that below 0°C, ΔT approaches a constant value which is the value for pure argon. At about 60°C, ΔT is seen to go through a shallow minimum, rising from that on upward, presumably all the way into the region of the high pressure Hg arc.

The circles in Fig. 3 are for ac, the crosses for dc operation. It is seen that the dc points lie about 1° higher than the ac points.

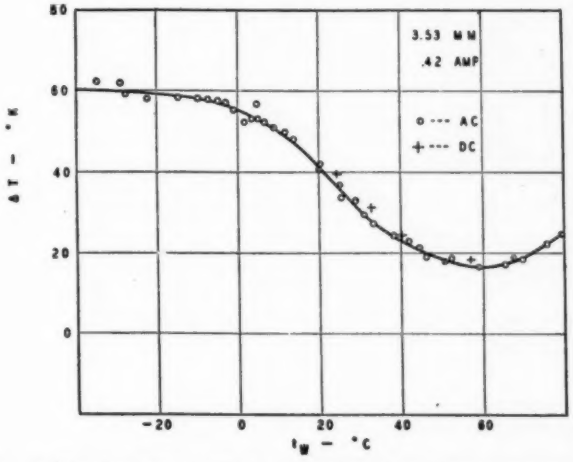


FIG. 3. Heating effect ΔT as a function of wall temperature t_w.

[3] Except in one special case, the discharge tubes were essentially identical to the 40-watt fluorescent lamp.

Measurements of ΔT as a function of argon pressure for an ac discharge in another tube are shown in Fig. 4. The heating effect is seen to increase substantially linearly with the pressure.

The average of all measurements with the two tubes, for 3.5 mm argon, 42°C wall temperature, and 0.42 amp ac gives the value of 22.4° for ΔT. The corresponding value for dc is taken from Fig. 3 as 23.3°.

ELIMINATION OF END EFFECTS

The above value of $\Delta T = 23.3°$ is for the entire tube. If it is desired to obtain the heating effect in the positive column alone, it is necessary to eliminate the end effects. This has been done by measuring ΔT for a tube similar to the others but relatively short.

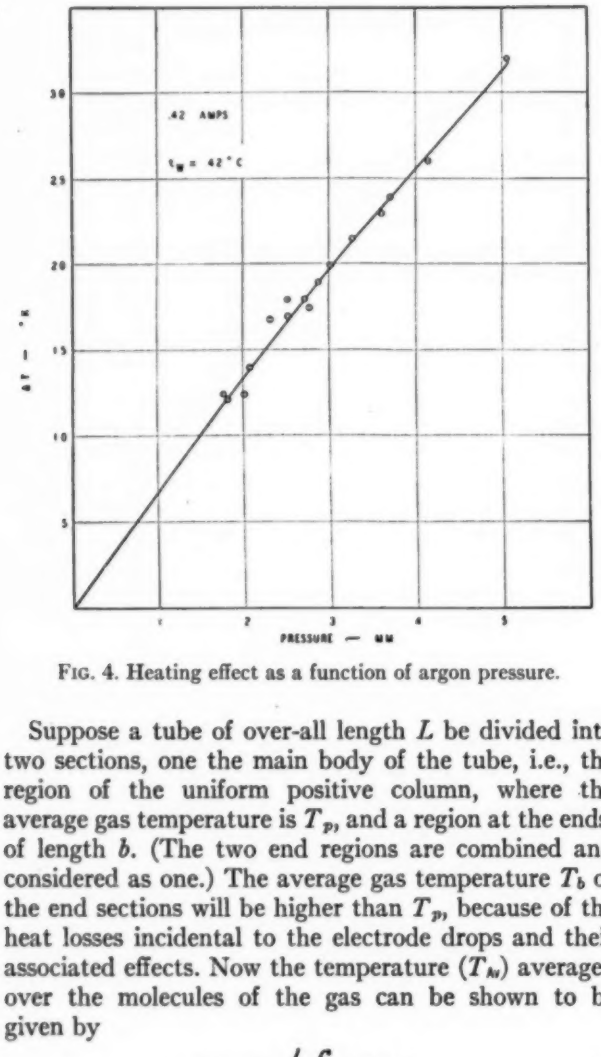


FIG. 4. Heating effect as a function of argon pressure.

Suppose a tube of over-all length L be divided into two sections, one the main body of the tube, i.e., the region of the uniform positive column, where the average gas temperature is T_p, and a region at the ends, of length b. (The two end regions are combined and considered as one.) The average gas temperature T_b of the end sections will be higher than T_p, because of the heat losses incidental to the electrode drops and their associated effects. Now the temperature (T_h) averaged over the molecules of the gas can be shown to be given by

$$T_h = V \Big/ \int_V (dV/T), \tag{5}$$

where T is the absolute temperature of the gas in the element of volume dV, and the integral is taken over the whole volume V. Applied to a tube of uniform cross section, Eq. (5) reduces to

$$T_h = L/[(L-b)/T_p + (b/T_b)]. \tag{6}$$

The use of two lengths of tube leads to two equations of the form of Eq. (6). Since there will be only minor differences of density, the end and middle temperatures will be approximately the same in both cases. The two equations may be treated as simultaneous, and elimination of b and T_b leads to

$$(L-l)/T_p = (L/T_L) - (l/T_l), \tag{7}$$

where L is for the long tube and l is for the short one. With ac operation, for $L = 118$ cm and $l = 16.5$ cm, the values $T_L = 337.4°$ and $T_l = 349°$K were found. Then Eq. (7) yields $T_p = 335.6°$ or $\Delta T = 20.6°$C.

Inasmuch as the operating voltages are observed to be closely the same at a given current for ac and dc discharges of equal length, it will be assumed that the over-all end effects are the same for both cases. Since the uncorrected values of ΔT differed by 1°, it will therefore be assumed that ΔT for the dc case is 21.6°C.

THE TEMPERATURE DISTRIBUTION

It is now desired to infer the actual temperature distribution corresponding to the above value of ΔT of 21.6°. To do this, the distribution of heat production and the conduction of heat to the walls must be considered. It will be assumed that to a first approximation the heat development in any element of volume will be proportional to the electron concentration n_e and that the radial variation of n_e will be parabolic. Actually, n_e will follow more nearly a bessel function, but the difference is considered negligible for the present purposes. Then, if R be the rate of heat production per cm³ in any element of volume, and R_0 the rate at the axis, it follows that

$$R = R_0(1 - r^2/a^2), \tag{9}$$

where r is the distance from the axis and a is the lamp radius (1.8 cm). If K be the coefficient of heat conduction, the differential equation for T, the absolute gas temperature (at any point distant r from the axis) will be

$$(d^2T/dr^2) + (1/r)(dT/dr) + (R_0/K)(1 - r^2/a^2) = 0. \tag{10}$$

Solution of this equation by the method of series gives

$$T = (T/a) + \tfrac{3}{16}(R_0a^2/K)[1 - (4/3)(r^2/a^2) + \tfrac{1}{3}(r^4/a^4)]. \tag{11}$$

This function of r, for a number of values of R_0/K and for $T_a = 315°$K (42°C), is plotted in Fig. 5, with r/a as abscissa.

For the case of cylindrical symmetry, Eq. (5) takes the form,

$$T_h = \frac{\pi a^2}{\int_0^a (2\pi r dr/T)} = \frac{1}{2\int_0^1 [(r/a)/T] d(r/a)}. \tag{12}$$

Using the curves of Fig. 5, the quantities $(r/a)/T$ were plotted against (r/a) and the integrations performed graphically. T_{Av} for the four values of R_0/K was then evaluated and the results plotted in Fig. 6. The measured T_{Av} of 336.6°K is then seen to correspond to a value of R_0/K of 83.5. This value of R_0/K when substituted in Eq. (11) gives an axial temperature of 366°K, 51° above the wall. This means that the argon density at the axis will be 86 percent of that at the wall. The Hg density at the axis will probably be between 86 percent and 93 percent of that at the wall, depending on the extent to which thermal transpiration is present. Any exact theory of the positive column must take into account these departures from uniform density.

The value of K for argon at the average operating temperature of about 337°K as found by extrapolating Euken's experimental values[4] is 4.5×10^{-5} cal sec^{-1} cm^{-1} deg^{-1}. This gives the value of R_0 then as $83.5\times4.5\times10^{-5}=3.76\times10^{-3}$ cal cm^{-3} sec^{-1}. Since for the parabolic distribution of R across the tube the average value of R will be half the maximum, the total heat production per cm of length will be $\pi a^2R_0/2=1.9\times10^{-2}$ cal sec^{-1} $=0.080$ watt. Multiplication by the length of the positive column (assumed to be about 112 cm) gives 8.9 watts as the elastic energy loss in this uniform column. The expected elastic loss will now be calculated, for comparison, from the known electron concentration, electron temperature, and cross sections for elastic collision.

COMPUTED LOSS

Equations for elastic loss,[5] given in the literature on gaseous discharges, ordinarily are applicable only to cases in which the mean free path of the electrons is independent of their velocity. The rare gases, many of the common gases, and mercury, cadmium, and zinc vapor, all have a marked variation[6] of electronic free path with velocity. In the case of argon, substitution of the free path value computed from the viscosity into the conventional equation[5] for elastic loss may result in errors as large as a factor of three. This difficulty arises because the equations for diffusion, mobility, and elastic loss have different powers of the electron velocity in them. Thus a different weighting must be given to the mean free path L_e for a particular velocity c, depending on the process for which one wishes to calculate an effective free path. This is evident from Table I. The curves in Fig. 7 give computed values of the effective free paths, for use in the equations[5] derived for constant mean free path, for maxwellian distributions of electron velocity in 3.5 mm (Hg) of argon plus

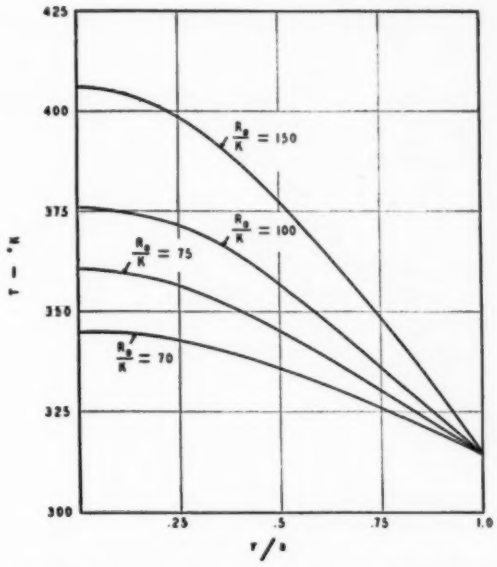


FIG. 5. Gas temperature as a function of fractional radial distance (r/a) for different values of the parameter R_0/K. R_0 = heat production per cm³ at the axis; K = heat conductivity.

0.006 mm of mercury vapor (pressures reduced to terms of 0°C).

Various authors state that one must take into account the Ramsauer effect in computing elastic loss, but apparently no one has actually treated mean free path as a variable in the integration of elastic losses over the velocity distribution of the electrons. A method of computing elastic loss which seems reliable will now be worked out in detail.

For electrons of a particular velocity c, the average fraction of the energy of each electron that is lost per collision is[7] $(2m/M)(1-E_1/E_2)$, where m = mass of elec-

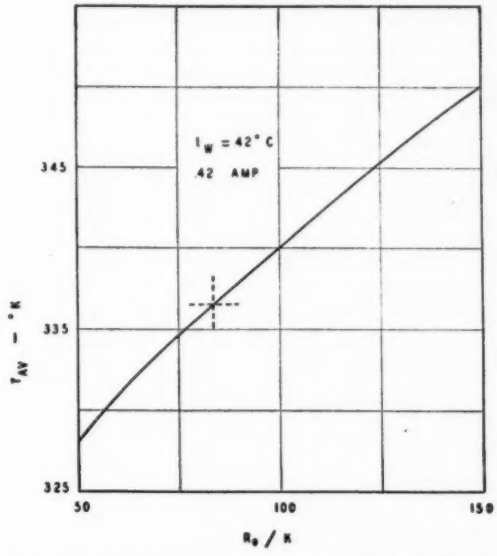


FIG. 6. Average gas temperature as a function of the parameter R_0/K. Cross represents value of R_0/K corresponding to measured T_{Av}.

[4] Values of K used are those of Eucken (see I. Amdur, J. Chem. Phys. **16**, 190 (1948)) extrapolated somewhat into the temperature region here involved. The radial variation of K for the case considered is about 9 percent; this will be neglected here.

[5] See, for example, L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley and Sons, Inc., New York, 1939), p. 590.

[6] R. B. Brode, Revs. Modern Phys. **5**, 257 (1933).

[7] K. T. Compton, Phys. Rev. **22**, 333 (1923).

TABLE I. Averaging procedure to be used when mean free path varies with electron velocity.

Desired result	Diffusion coefficient	Mobility	Elastic loss
Quantity averaged	L_e	L_e	$1/L_e$
Weighting factor	c	$1/c$	c^3

tron, M=mass of atom, E_1=mean energy of atoms, and E_2=electron energy. Then the average energy loss per collision is $(m^2c^2/M)[1-(3kT_g/mc^2)]$, where k=Boltzmann constant, and T_g=gas temperature. Since the number of collisions per electron per second is c/L_e, the power expenditure per second for an electron of velocity c is $(m^2c^3/ML_e)[1-(3kT_g/mc^2)]$. For all of the electrons in a velocity distribution, the average power spent per electron per second is

$$\frac{m^2}{M}\int_0^\infty \frac{c^3}{L_e}\left(1-\frac{3kT_g}{mc^2}\right)f(c)dc, \qquad (13)$$

where $f(c)dc$ is the fraction of the electrons with velocities between c and $c+dc$.

Tabular integrations of the expression derived in the preceding paragraph were made for various maxwellian distributions of electron velocity, for mercury and for argon. The results are shown by Fig. 8. The probability of collision *vs* electron velocity for argon was taken from the curve given by Brode.[8] This curve is based on data taken by Ramsauer and Kollath. Since the latter give a much smaller minimum value than that found by Normand,[9] there is some uncertainty as to the true values of the elastic loss for low energy electrons (T_e's<7000°K).

The values used for the probability of collision *vs* the electron velocity for mercury were taken from Brode's original data,[10] but were multiplied[10] by 1.10 in order to put them in terms of the mercury vapor pressures given in the *International Critical Tables*. The great difference in the slope of the two curves in Fig. 8 reflects the fact that, for mercury, the probability of collision decreases with increasing electron velocity, but with argon the opposite is true.

The curves of Fig. 8 furnish a means of computing the elastic loss for any mercury-argon discharges for which the electron temperature and concentration are known, provided the electron velocity distribution follows the maxwellian distribution law. Results for discharges in 3.2 mm pure argon (pressure at 0°C) mixed with mercury vapor are given in Tables II and III. Electron temperatures were measured with

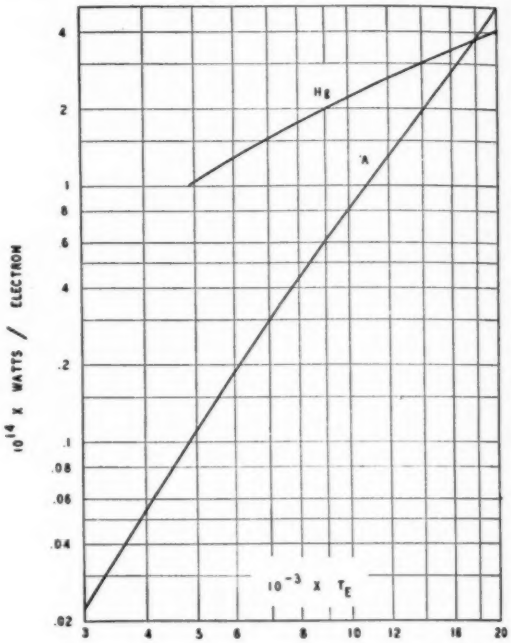


FIG. 8. Elastic loss *vs* electron temperature in mercury and in argon for 1 mm (Hg) pressure at 0°C.

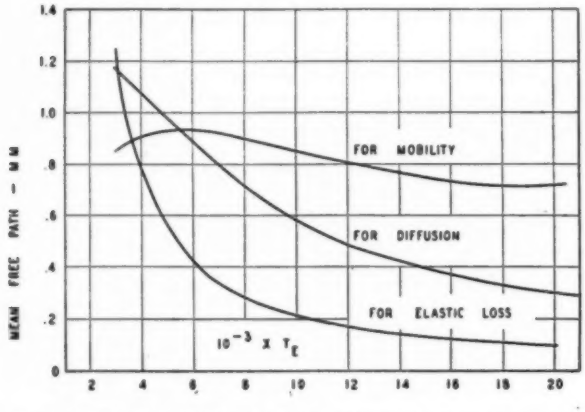


FIG. 7. Effective electronic free paths *vs* electron temperature in 3.5 mm (Hg) of argon plus 6 microns of mercury (pressures reduced to 0°C).

wire probes centered at the axis of the positive column. Maxwellian velocity distribution was indicated by the linearity of the plot of the logarithm of the electron current to the probe *versus* probe voltage over the measurable range of electron current (i.e., for a range 0 to 9–11 volts from space potential). Electron distribution over the cross section (πa^2) of the discharge was assumed to follow a bessel function[11] with N_e, the number of electrons per centimeter length of column, given by the equation,

$$N_e=2\pi a^2(J_1(\mu)/\mu)n_0.$$

The root μ of the first-order bessel function J_1 was determined from the electron concentrations at the sheath edge (n_w) and at the axis (n_0), using the relation

$$n_w/n_0=J_0(\mu).$$

[8] R. B. Brode, Revs. Modern Phys. **5**, 263 (1933).
[9] C. E. Normand, Phys. Rev. **35**, 1222 (1930).
[10] R. B. Brode, Proc. Roy. Soc. (London) **A125**, 136 (1929).
[11] V. Fabrikant, Compt. rend. acad. sci. U. R. S. S. **XXIV**, 531 (1939).

TABLE II. Elastic loss *vs* mercury pressure in the positive column of a 0.42-amp dc discharge in mercury vapor mixed with 3.5 mm argon (tube diameter 36 mm).

Bath Temp. °C	Hg Pres. μ	T_e °K	$N_e \times 10^{-12}$	ΔT °C	P_V-watts/cm from $T_e - N_e$	P_V-watts/cm from ΔT
17°	0.95	15,000°	2.5	42.7	0.19	0.16
30°	2.8	13,300°	2.2	29.4	0.12	0.11
40°	6.1	11,300°	2.3	21.8	0.083	0.082
50°	12.7	10,100°	2.6_5	16.8	0.069	0.064
60°	25.2	9,800°	2.6_5	15.8	0.064	0.061

The energy loss P_V in watts per cm column was obtained, then, using the product of N_e and the loss per electron at the given electron temperature, from Fig. 8. It should be noted that the discharges described by these values of $T_e - N_e$ showed no striations in the range of oscilloscope sensitivity (i.e., up to 2 mc) and no obvious effects of impurities. The presence of oscillations or of small amounts of certain molecular impurities would change the form of the distribution function $f(c)$ in Eq. (13) and thus make inapplicable the curves of Fig. 8.

For a comparison of the results obtained by the two methods, the measured values of ΔT were converted to watts per centimeter, and are listed in Tables II and III. The importance of T_e is shown by these data. The weighting factor c^3 and the rapid decrease in the cross section for low electron energies combine to produce a marked falling off in the heating effect for low speed electrons. Thus, for an electron temperature decrease from 15,000° to 9800°, there was a threefold decrease in the electron energy lost in elastic collisions.

TABLE III. Elastic loss *vs* discharge current in the positive column of a dc discharge in 7μ of mercury mixed with 3.5 mm of argon (tube diameter 36 mm).

Current amp	T_e °K	$N_e \times 10^{-12}$	ΔT °C	P_V-watts/cm from $T_e - N_e$	P_V-watts/cm from ΔT
0.2	11,900°	1.1_5	13.1	0.047	0.047
0.4	11,400°	2.2_5	21.6	0.081	0.082
0.6	10,800°	3.3	27.1	0.104	0.107

DISCUSSION

The good agreement obtained between the elastic losses determined by the two methods is ascribed to reliable values of T_e and N_e made possible by the absence of striations, purity of the gas in all cases, and the proper use in the calculations of the Ramsauer cross-section curve. The course of the Ramsauer curve is not certain for electron energies below about one volt. If similar measurements could be made on an argon discharge where T_e were very low, the method might be used to furnish more information about the low energy end of the Ramsauer curve. Such a discharge might be one using Na or Cs with the argon, instead of Hg. The experiment would, of course, have to be done at an elevated tube-wall temperature.

A further use of ΔT measurements might be in the study of striated discharges.[12] Here it is difficult at present to obtain an exact knowledge of the distribution of T_e and N_e throughout the striation. But any tentative distribution could be tested for plausibility by spatial integration to see whether it gave the measured ΔT.

We are indebted to Dorothy Kindinger Williams for assistance in making the gas temperature measurements.

[12] W. Pupp, Z. Physik **67**, 237 (1931).

Pulse-Annealing for the Study of Relaxation Processes in Solids*

W. E. Parkins, G. J. Dienes, and F. W. Brown†
Atomic Energy Research Department, North American Aviation, Inc., Downey, California
(Received January 22, 1951)

A new experimental technique is presented for studying rate processes in solids. The method, termed "pulse-annealing," involves direct heating of a very small sample to a chosen annealing temperature for a short period, followed by rapid cooling to some fixed temperature for an in-place physical property measurement. The method is a very flexible one, as the above cycle may be repeated to obtain any desired time-temperature history. General experimental techniques are described and specific application to graphite and metals noted.

The annealing process is examined theoretically. A method is described for deriving the activation energy of the fundamental relaxation process from the isothermal annealing curves of a physical property, the physical property being used as an index of the annealing process. It is shown that the activation energy is derivable without knowledge of the relation between the measured physical property and the fundamental kinetic variable. For further kinetic analysis (determination of rate constants and orders of reaction), however, it is necessary to know this relation. The analysis of isothermal and tempering experiments is compared.

INTRODUCTION

In the study of chemical rate processes it has been possible to determine constants characteristic of a given reaction by measurement of the reaction rate at different temperatures. Constants which can be assigned are the activation energy, rate constant, and order of reaction. In a similar fashion it is possible to study the changes which take place under certain conditions within the lattice of a solid. Types of lattice processes which lend themselves to such analysis are healing of lattice imperfections, allotropic changes, order-disorder transformations, and precipitation. In determining the constants associated with these lattice processes, much can be learned from measurement of the rate of change of any given physical property as a function of heat treatment of the solid in question. In the theoretical section of the present paper, some of the possibilities of this approach are discussed.

In the performance of heat treatment of a solid and measurement of the associated changes in one or more physical properties, two methods have been commonly used. One method consists of annealing the sample by means of a chosen heat treatment in a furnace, cooling to remove the sample from the furnace, and finally making a physical property determination at some fixed temperature. By repeating this process using different heat treatments, the data necessary for analysis may be obtained. This method suffers from two disadvantages. It is slow because of the large number of separate cycles required in annealing and measuring the sample. In addition, the long heating period in reaching each desired annealing temperature requires a long period at an accurately controlled annealing temperature to give interpretable results.

The second method which has been used in heat treating and measuring associated physical property changes is an in-place measurement on the sample in the furnace and at the annealing temperature. This method permits rapid experimentation but suffers another disadvantage. Since the measurements must be made at different annealing temperatures, it is impossible to interpret the results unless the temperature dependence of the physical property is known for the various lattice configurations assumed by the solid during the course of the measurements.

A new method has been developed which is relatively fast and requires no knowledge of temperature dependence of the physical properties. This method has been termed "pulse-annealing." It involves direct heating of a very small sample to a chosen annealing temperature for a short period, followed by rapid cooling to a fixed temperature for an in-place physical property measurement. This cycle is repeated to obtain data for any desired time-temperature history. To date, the method has only been applied to materials which can be heated by the passage of an electric current, although it might well be adapted to insulating materials if dielectric heating were employed.

Apparatus has been built and successfully used at the Atomic Energy Research Department laboratories of North American Aviation, Inc., to measure the changes during pulse-annealing in the Hall coefficient, magnetoresistivity coefficient, thermal conductivity, thermoelectric power, and electric resistivity. Typical results of measurements of the above galvano-magnetic properties have been described by Donoghue and Eatherly.[1] The apparatus and some results for measurements of the above thermal properties have been reported by Rasor and Parkins.[2] To illustrate the type of problems encountered in pulse-annealing studies, the present article will be confined to apparatus and results for changes in electric resistivity. In addition, a

* This work supported by Contract AT-11-1-GEN-8 of the AEC.

† Now at Naval Ordnance Test Station, Inyokern, China Lake, California.

[1] J. J. Donoghue and W. P. Eatherly, Rev. Sci. Instr. (to be published).

[2] N. S. Rasor and W. E. Parkins, Phys. Rev. 78, 638 (1950).

general discussion will be given of the pulse-annealing method, its uses, and its limitations.

PULSE-ANNEALING APPARATUS

Specimens of the solid to be studied are prepared in the form of narrow ribbons which can be heated by the passage of 60-cycle ac current from a variable voltage supply. The ends of the specimen are clamped to heavy current leads which are cooled by a circulating fluid whose temperature is near that at which the physical property determinations are to be made. The apparatus must be arranged so that the physical property measurement are made only on some central region of the specimen which receives an identical temperature history during any annealing pulse. The specimen and its directly associated apparatus are surrounded by a vacuum or some inert atmosphere which can have no deleterious effect on the specimen during the annealing pulse and cannot interfere with the subsequent physical property measurement. It is highly desirable to arrange the apparatus so that an entire pulse-annealing study can be performed on a single specimen without requiring any modification inside the vacuum (or inert atmosphere) envelope.

Figure 1 shows an arrangement which has been successfully used in connection with studies on graphite. A specimen, 1 inch long, 0.10 inch wide, and 0.030 inch thick, was held between water-cooled copper clamps. These clamps carried the ac heating current during an annealing pulse and a very small dc current during the subsequent resistance measurement. The resistance of the center region of the specimen was determined from the emf measured between two probes in contact with the specimen and 0.080 inch apart. During annealing periods to temperatures as high as 2200°C, the region between these probes never acquired temperature differences larger than 5°C. The total specimen length exposed between the clamps was approximately 0.75 inch. The specimen unit was mounted on a block of fired lavite and a guide provided to permit the expansion of the graphite during annealing pulses. One clamp was made sufficiently wide to shield the lavite during the annealing pulse in order to minimize back radiation. This assembly was supported on the copper tubes which served to carry electric current and water for cooling to the ends of the specimen. During operation the entire unit was enclosed in a vacuum bell jar normally operated at about 10^{-5} mm Hg.

The probes which touch the specimen near its center must not permit variation in the effective position of the electric contact during the entire annealing run. One scheme, which was usually successful, made use of 0.005-inch diameter tungsten wire probes, sharpened by etching, and simply held against the graphite surface by spring tension. A more reliable arrangement, shown in Fig. 1, consisted of force fitting 0.008-inch diameter stubs into holes drilled through the graphite. The wire stubs were approximately 0.050 inch long and made

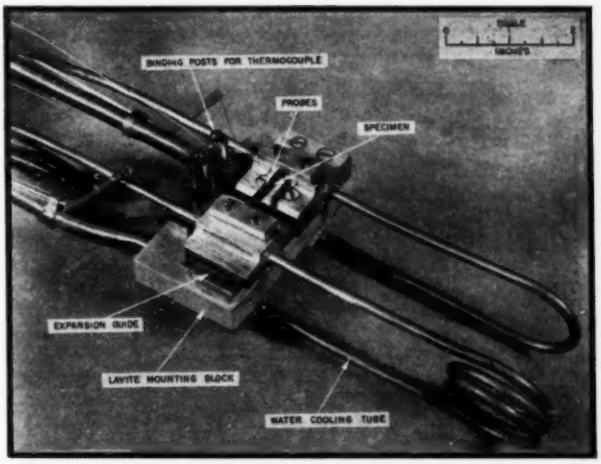


FIG. 1. Graphite specimen mounted in holder for electric resistivity measurements during pulse-annealing.

of W, Ta, or Mo. Electrical contact was made by means of smaller diameter W leads arranged to press lightly against the stubs, while a small loop at the end of each lead prevented lateral slipping. During annealing to high temperatures a fraction of the stub underwent conversion to the metal carbide, but this introduced a negligible resistance change.

For measurement of the central specimen temperature during an annealing pulse, a Pt to Pt–10 percent Rh thermocouple was used at temperatures below its melting point. An optical pyrometer was used at higher temperatures and provided a check on the thermocouple over the range from approximately 800°C to 1700°C. The thermocouple was made by butt welding 0.005-inch diameter Pt and Pt-Rh wires in a specially made jig.[3] The thermocouple wire was then threaded through a 0.006-inch diameter hole drilled through the specimen just outside the probe region. When the thermocouple junction reached the midpoint of this hole, the wires were bent to prevent further motion and a tiny amount of sodium silicate cement was applied at the thermocouple hole to improve thermal contact between the thermocouple wire and the graphite. In the optical pyrometer range it was possible to obtain several pyrometer temperature determinations during a single annealing pulse. The average of these gave a temperature whose reliability was comparable to that obtained using the thermocouple.

The electric circuit for controlled heating of the specimen as well as for accurate resistance measurement is shown in Fig. 2. For heating, power from a voltage regulator is fed to a variac and transformer arrangement. The output of the transformers may be connected across the specimen in series with a load resistor, or across the load resistor alone. The purpose of the load resistor is to overcome instability in the circuit which may be produced by changes in specimen

[3] E. F. Hammel, MDDC 776 or LADC 393.

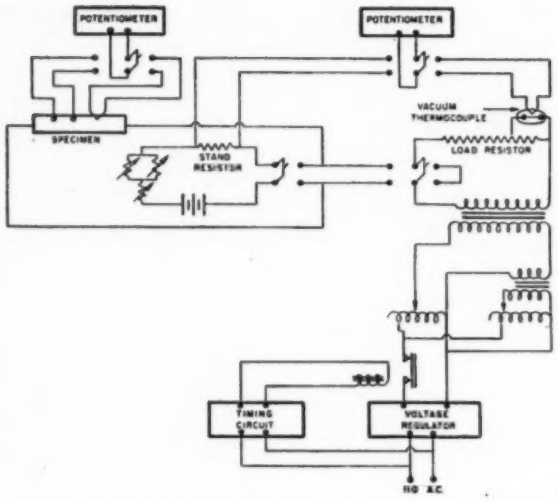


FIG. 2. Circuit for electric resistivity pulse-annealing apparatus.

resistance during heating and to serve as a means for accurately monitoring the transformer output. A vacuum thermocouple connected across a portion of the load resistor gives a precise measurement of the amount of load resistor current. It is sometimes desirable to adjust the variacs for heating power on the basis of specimen current. This can be done when the load resistor is connected in series with the specimen and the variacs are adjusted to give a particular vacuum thermocouple reading. Such a procedure is useful, for example, when the specimen is being pulsed to successively higher temperatures. In this case a running plot of specimen temperature as a function of vacuum thermocouple reading is maintained during the course of the measurements.

In other situations it is sometimes desirable to adjust the variacs to a given vacuum thermocouple reading when the specimen is not in the circuit. This procedure has proved useful when the specimen is being pulsed to successively higher temperatures and when it is necessary to keep the time for specimen heating as short as possible. After setting the variacs to some particular vacuum thermocouple reading, the switch is thrown which connects the specimen in series with the load resistor and produces the annealing pulse without further variac adjustment. In this case the specimen's annealing temperature must be plotted against the vacuum thermocouple reading used in determining the variac setting, thereby providing a running record for predicting later variac settings.

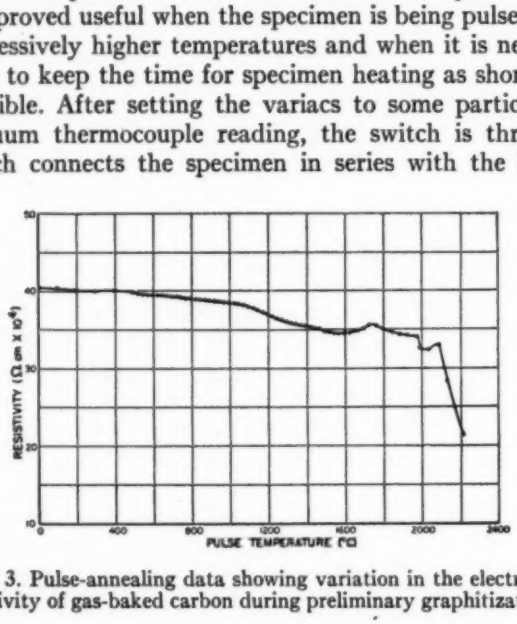


FIG. 3. Pulse-annealing data showing variation in the electric resistivity of gas-baked carbon during preliminary graphitization.

Usually the few seconds required for making the variac setting during actual specimen heating are of no disadvantage. Under such circumstances, if a specimen thermocouple is available, it is a simple matter to advance the variacs to the proper position based on the specimen thermocouple reading. It is always very important to avoid overshooting the desired specimen annealing temperature (specimen temperature implying that of the central region subjected to resistance measurements).

To avoid resistance variations in the load resistor due to its own temperature changes, the resistor is made of nichrome ribbon submerged in a bath of propylene glycol. The bath is cooled, in turn, by a submerged coil of copper tubing in which tap water is circulated. Potentiometers, which aid in the specimen resistance measurement after each annealing pulse, are appropriately switched to measure the specimen thermocouple and vacuum thermocouple emf's for the annealing pulses.

If the variac setting is made before the heating pulse and no further readjustment is necessary, the approach to equilibrium annealing temperature is very nearly exponential. Likewise, the specimen cooling from the annealing temperature when the heating power is switched off is nearly exponential. Ideally, the temperature-time function during any annealing pulse would be a square wave. Knowing the time constant for the exponential temperature rise, it is possible to correct for annealing which occurs during the approach to equilibrium temperature, although this complicates the analytical treatment of the experimental results. It is usually possible to neglect any annealing which occurs during specimen cooling. Also, it is usually true that by using heating pulses of one minute or more duration, it is possible to neglect all annealing except that taking place at the equilibrium pulse temperature. Approximate analysis indicates that with a time constant of the order of one second, the timing error for a one-minute pulse is only a few percent and is, of course, even more negligible for longer pulses. Once an approximate kinetic expression is available for the relaxation process being studied, the timing correction is easily calculable and the data can be refined accordingly.

Perhaps the most important feature of the technique described above is the accurate and rapid measurement of changes in physical properties up to very high temperatures. This is of paramount importance for such a material as graphite. Most of the specific problems on graphite, to which the technique has been applied successfully, are of a classified nature. As an example

of the type of results which can be obtained, Fig. 3 is presented. Measurements were made on a specimen of gas-baked carbon, during pulse-annealing to 2215°C. The annealing schedule consisted of one-minute heating periods at successively higher temperatures, separated by about 25°C, in the region above 500°C. The electric resistivity was measured at 20°C after each annealing pulse and is plotted in Fig. 3 as a function of the annealing temperature. The absolute values of the resistivity depend on initial dimensional measurements of the region of specimen between the probes and, hence, are not as accurate as the relative values.

The gas-baked carbon was originally made by calcining a mixture of coke and pitch at a temperature of about 1000°C. The pulse-annealing procedure simulated the conditions which would take place during the early stages of graphitization when the gas-baked carbon was heated in an electric furnace to temperatures in the neighborhood of 2500°C. The variation in electric resistivity measured during the pulse-annealing is, therefore, characteristic of that taking place during the preliminary graphitization process. Unfortunately, it was impossible to carry the graphitization to completion because of the appreciable carbon evaporation which sets in at about 2200°C. Most of the resistivity changes measured, including both increases and decreases, are presumably due to the diffusion of various impurities out of the sample. The sharp drop in resistivity beginning at about 2100°C is largely a result of growth of the graphite crystals. As this process is carried to completion, the resistivity is known to drop to a value approximately one-half of that obtained after the 2215°C annealing pulse.

MODIFICATIONS FOR PULSE-ANNEALING OF METALS

Pulse-annealing studies can be performed on metals in a similar manner to that just described for graphite. Some modifications, however, are necessary. Metal specimens should be made in the form of small wires or of very thin, narrow strips. This smaller cross section increases the electrical and thermal resistance, thereby simplifying the heating problem. Also, it aids in creating a sufficiently flat temperature distribution over the central region of the specimen. In cases where the temperature uniformity is critical, it is advisable to connect the ends of the specimen to thin ribbons of some other metal which, in turn, are attached to the cooled copper clamps. Such a pulse-annealing arrangement is shown in cross section in Fig. 4. The same apparatus is illustrated more fully in Fig. 5.

To attach a thermocouple for determining the central specimen temperature, a spot-welding technique has been most successful. The two thermocouple wires are welded, one at a time, to the specimen surface, thereby making a junction and attaching it simultaneously. The thermocouple wires are of small diameter, about 0.003 inch, to prevent any appreciable

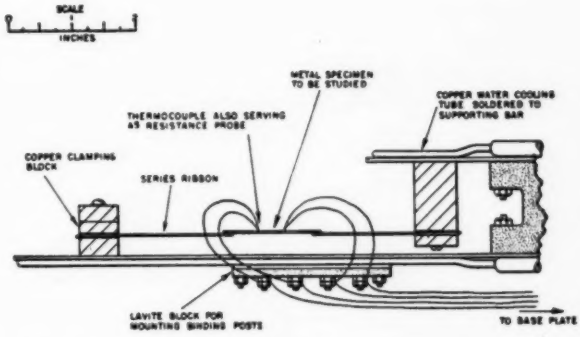


FIG. 4. Apparatus for mounting metal specimen for electric resistivity measurements during pulse-annealing.

cooling effect during the annealing pulse. To spot-weld, the fine thermocouple wire is brought in contact with the specimen surface, and then a capacitor is discharged through a series circuit which contains the thermocouple wire and specimen. Satisfactory welds can be obtained with practically any material with proper choice of capacitor size and charging voltage.

To attach probes to aid in the resistance measurement, a similar technique to that just described for thermocouples may be employed. Actually, it has been found practical to attach one thermocouple junction at the site of each of the two probes and to use the thermocouple leads to measure temperature during the annealing pulse and the central specimen emf during the subsequent resistance determination. The two central thermocouples are useful in giving some check on temperature uniformity. To measure the emf for the resistance determination, two like wires, one from each of the junctions, are employed.

It is possible to make measurements with sharpened probes held against the specimen surface by spring tension, but this practice is not recommended. It is even less desirable to depend on such an arrangement for the thermocouples. The spot-welding technique described above has proved very satisfactory. The volume of metal on the specimen surface which is heated during the instant of welding is so small that

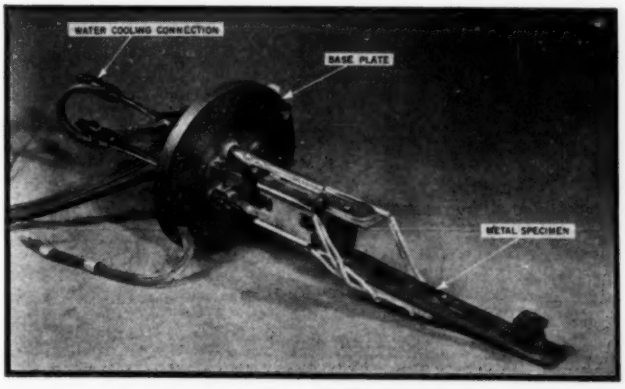


FIG. 5. Photograph of the pulse-annealing apparatus for metal specimens.

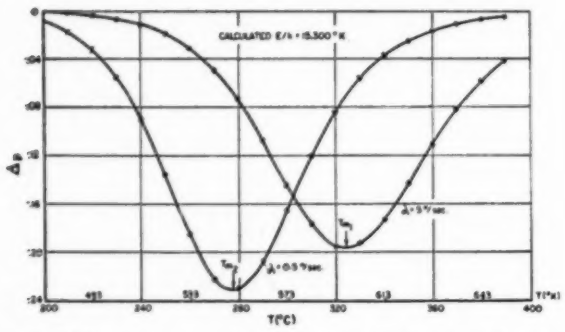


FIG. 6. Difference curves for p from tempering experiments at two different rates of heating.

there is no detectable effect on the properties of the specimen to be studied during the pulse-annealing schedule. In any case, the problem of the use of satisfactory probes and thermocouples for metal specimens is very simple compared with that for graphite. This is because the welding technique cannot be used with graphite and because the annealing temperature range of interest is much lower for most metals.

The apparatus shown in Figs. 4 and 5 was designed to permit the use of ribbons of adjustable length, connected in series to the specimen. One ribbon is fastened to each end of the specimen to improve its temperature uniformity during annealing. The ribbons are clamped to copper blocks which are fastened, in turn, to water cooled supporting bars. These supporting bars are slotted to provide adjustable locations for the clamping blocks. In studies on AuCu alloys, a molybdenum ribbon, 0.005 inch by 0.125 inch by 3 inches, was connected to each end of a 1 inch long specimen. The cross-sectional dimensions of the specimen were similar to those of the ribbons. To connect the molybdenum and AuCu without introducing additional heat capacity, the two materials were silver soldered together by means of a thin sheet of silver solder spot-welded between the speciment and the ribbon. This technique had the additional advantage of avoiding any heating of the central region of the specimen during assembly.

The entire apparatus, shown in Fig. 4 and 5, was inserted into a bell jar and a vacuum seal was made at the base plate. Experiments have been conducted using an inert gas instead of vacuum. The gas has the advantage, important in some cases, of shortening the time required for specimen cooling at the end of each heating pulse. With the AuCu alloy, argon was successfully used for this purpose, whereas helium was found unsatisfactory since it caused the specimen material to become brittle. Other aspects of the experimental procedure used with metal specimens were identical to those previously described for graphite.

The technique described has been applied successfully to the study of the kinetics of ordering in the 50–50 atomic percent AuCu alloy. These results are fully described elsewhere.[4]

[4] G. J. Dienes, J. Appl. Phys. **22**, 1020 (1951), companion paper.

GENERAL THEORETICAL CONSIDERATIONS

The experimental technique, described in previous sections of this paper, is a very flexible one in that almost any desirable temperature-time history during pulse-annealing can be realized. It is pertinent to inquire, therefore, as to the type of information derivable from different kinds of experiments. The main purpose of this section is to discuss the theory of solid-state annealing or relaxation processes in general and the analysis of some typical experiments often encountered in practice.

Let us assume that changes in physical properties are the result of changes in the number, n, of some fundamental lattice disturbance. Further, we shall assume that n varies with annealing time according to the laws of simple, single-stage rate processes; i.e., a change in n is governed by a chemical rate equation

$$dn/dt = -K_0 n^\gamma e^{-E/kT}, \tag{1}$$

where K_0 = specific rate constant, γ = order of reaction, and E = activation energy. Equation (1) is evidently not applicable to multistage processes. In these cases our discussion is applicable only to the individual steps constituting the over-all process. Complete description of the annealing process involves the determination of K_0, γ, and E, as well as the dependence of the measured physical property, which is employed to follow the reaction, on the fundamental kinetic variable n.

1. Phenomenological Determination of the Activation Energy

The most important characteristic constant is the activation energy, E, of the annealing process. In this section it will be shown how E can be derived from experimental data with a minimum of detailed knowledge of the relaxation process and a minimum of restrictive theoretical assumptions.

Two types of heat treatments have been commonly used; namely, isothermal heat treatment and tempering treatment. During an isothermal heat treatment a physical property, p, is measured as a function of time at a constant annealing temperature. For analysis, the physical property must be referred to a fixed temperature. Consequently, if p is known as a function of T for each state of the material during annealing, measurements at annealing temperatures are satisfactory. We shall take this temperature coefficient as unknown and insist that p itself be measured at some fixed temperature independent of the annealing temperature. In a tempering-type experiment the specimen is subjected to a fixed duration of anneal (usually short) at successively higher temperatures. Again, we assume that p is measured at some fixed temperature. The theory will show that E is unambiguously derivable from isothermal experiments, while tempering experiments can be analyzed only approximately without introducing restrictive assumptions. Isothermal experiments will be covered first.

The physical property, p, which is used as an index of the reaction, must be connected to n at the fixed temperature of measurement (which is not, in general, equal to the annealing temperature) by a functional relation

$$p=g(n). \tag{2}$$

The only restriction we shall impose on $g(n)$ is that it shall be a single-valued function. The general theory will be given for any $\gamma\neq 1$. Obvious modifications, which do not influence the results, are necessary for $\gamma=1$.

At a constant temperature, T, Eq. (1) can be integrated with respect to time, t, to give

$$n_0^{1-\gamma}-n^{1-\gamma}=K_0(1-\gamma)e^{-E/kT}t, \tag{3}$$

where n_0 is the value of n at $t=0$. Elimination of n by combining Eqs. (2) and (3) gives

$$[1/K_0(1-\gamma)][h(p_0)-h(p)]=te^{-E/kT}=u, \tag{4}$$

where p_0 is the value of p at $t=0$ corresponding to n_0, and h is an unspecified function.

If identical samples (same p_0) are now exposed to different temperatures for varying lenghts of time and the relaxation in p is measured as a function of time, the activation energy can be determined unambiguously without knowing the form of $h(p)$. This is accomplished by determining the times and temperatures resulting in a particular value of p, each experiment being run isothermally and always with the same value of p_0. For such an annealing treatment, u in Eq. (4) must be constant. Thus, if $p=c_1$, then

$$u=te^{-E/kT}=c_2 \tag{5}$$

or

$$\log_{10}t=\log_{10}c_2+(E/2.303kT). \tag{6}$$

From Eq. (6), the activation energy, E, is immediately calculable. The value of E should be independent of c_1. The analysis is only applicable to a state of single activation energy. If any relaxation is not describable by a constant E, then $\log_{10}t$ of Eq. (6) will not be linear when plotted against $1/T$, and E will not be independent of c_1. Thus, on the basis of the analysis presented here, the E value can be determined without any but the most general assumptions concerning the relaxation process. Furthermore, the validity of assigning a constant value of E to any given relaxation state is automatically judged by the applicability of inapplicability of Eq. (6). It is to be emphasized that this analysis can be applied to the annealing of any physical property as long as n is a single-valued function of p.

It may be pointed out here that from a purely mathematical standpoint a function $p=f(n)$ can always be constructed such that n itself obeys the kinetic rate equation, Eq. (1). This means that, in the absence of further knowledge about the process under investigation, the physical significance of n remains obscure. The phenomenological activation energy, however, has considerable physical reality for any single-stage process, since it is a constant of the process independent of the kinetic behavior of n and is always defined and determinable. This method has been used successfully in a number of instances for activation energy determination. An example is given in the accompanying paper[4] for ordering in AuCu.

As stated above, if the data do not conform to Eq. (6), then the process is not describable by a single activation energy and a distribution in E may have to be introduced leading to a considerable mathematical complexity. Vand[5] has done this on the basis of tempering experiments and has discussed approximate methods for evaluating the distribution function. This approach is complicated by the necessity of assuming an order for the process as well as a relation between p and n. To our knowledge, no clear-cut experiment has been presented in the literature for which a distribution in E is necessary to describe the results.

As will become evident later, it is very difficult to determine E from more complex heat treatment experiments unless further knowledge about the kinetics, and particularly the relation between p and n, is available. One approximate method, which we have used with some success, should be mentioned. Suppose that a typical tempering experiment has been run with the sample subjected to pulses of fixed duration at successively higher temperatures. Let the pulses be short enough and the successive temperatures at close enough intervals so that the temperature-time history may be approximated by a linear relation

$$T=T_0+\alpha t,$$

where α is the rate of heating. The difference curve, $\Delta p(=p_i-p_{i-1})$ *vs* T, will go through a minimum, as illustrated in Fig. 6 for a process of activation temperature $E/k=15{,}000$°K for two different rates of heating. The curves of Fig. 6 were calculated for stepwise annealing using 10°C temperature increments with 2- and 20-second pulses corresponding to heating rates α_1 and α_2, respectively. Time scales and heating rates are only illustrative and are not actual experimental schedules. It can be shown that to a good approximation the temperatures at the minima and the corresponding rates of heating are connected by the equation

$$(1/\alpha_1)T_{m_1}^2\exp(-E/kT_{m_1}) = (1/\alpha_2)T_{m_2}^2\exp(-E/kT_{m_2}), \tag{7}$$

where α_1, α_2 = rate of heating in degrees per second, T_{m_1}, T_{m_2} = temperatures of the minima in the differential pulse-annealing curves at rates of heating α_1, and α_2, respectively (in °K), and E = activation energy. The detailed derivation of Eq. (7) is given in the Appendix. From the curves of Fig. 6, an E/k of 15,300 is calculated. The advantage of this method is that it often permits a more rapid estimation of E than

[5] V. Vand, Proc. Roy. Soc. (London) **55**, 222 (1943).

does the isothermal procedure. It is an approximation, however, and requires an accurate determination of the minimum point in the differential curve, which may be experimentally difficult. It should also be pointed out that in tempering experiments errors in measurement are cumulative to a large extent, since every point p_i depends on the results of all previous points. In practice, Eq. (7) has been found to be quite reliable, as long as the heating schedule is well approximated by a linear temperature rise, and is a useful technique in giving a rapid but approximate value for E. When accurate activation energy values are desired, the technique of long time anneal at several fixed temperatures is advocated.

Another point which is pertinent to the phenomenological theory should be discussed. It may happen that the physical property, p, itself obeys a rate equation of the form

$$dp/dt = -K_0' p^\alpha e^{-E/kT}. \quad (8)$$

From a comparison of Eq. (8) and Eq. (1), it is easily shown that for p to obey this rate equation the connection between p and n must be of the form

$$p = cn^\beta$$

with the variables so adjusted that $p=0$ at $n=0$. Only in such cases are a phenomenological order and associated rate constant derivable.

In general, the connection between p and n must be determined independently before the kinetics of the process can be elucidated. Since the unspecified function, $h(p)$, in Eq. (4) contains the order of the reaction and its rate constant as well as the physical property dependence on n, these two effects cannot be separated without further independent information. By a purely phenomenological analysis, only the activation energy can be evaluated.

2. Kinetic Analysis

In this section kinetic analysis of various experiments will be presented with particular emphasis on a comparison of the two often used temperature-time schedules; namely, isothermal and tempering treatments. In the following it will be assumed that the functional relation between the physical property, p, and the kinetic variable, n, has been determined. The analysis which follows illustrates how K_0, γ, and E may be obtained from isothermal and from tempering experiments, where the measured changes in the physical property are a result of relaxation of a discrete activation state of energy E. We have carried out similar analyses showing how this same information can be obtained where the activation state has some form of continuous energy distribution. For example, methods of analyzing square and gaussian energy distributions have been worked out. None of these special cases is considered a likely possibility, and only the analysis for the discrete case is presented.

Isothermal rate curves are easily analyzed by means of standard techniques based on integrated rate equations for the order of reaction, γ, and the rate constant K, where $K=K_0e^{-E/kT}$. The activation energy, E, is then determined in the usual manner from the variation of the rate constant with temperature. For consistency of analysis, the activation energy determined in this way should agree closely with the value derived from a phenomenological analysis.

The importance of the order of the reaction should not be overlooked. A high order reaction is characterized by a rapidly decreasing reaction rate with increasing time. The point we wish to make is the following: It has been fashionable to regard solid-state rate processes as being of the first order. In terms of a first-order reaction, experimental curves characterized by higher orders would indicate, because of their rapid leveling off, that equilibrium reactions or a distribution of activation energies are involved. A careful scrutiny of the kinetics in terms of different possible values for γ is advocated in order to distinguish more clearly between single-step and multistage mechanisms. For such a purpose, we feel that the isothermal technique is the most reliable. The kinetics derived from isothermal experiments should be further substantiated by carrying out experiments at different values[4] of n_0.

A tempering experiment is defined here as one in which the specimen is subjected to temperature pulses of definite duration, with each pulse at a temperature higher than the previous pulse. We shall analyze tempering experiments where the pulse duration is kept constant. More complex annealing schedules can be treated in an analogous manner. Let the temperature of the ith pulse be denoted by T_i and the duration of the pulse by $\Delta t_i = t_i - t_{i-1}$. For any given pulse, the kinetic rate equation can be integrated to give

$$(1/n_i^{\gamma-1}) - (1/n_{i-1}^{\gamma-1}) = -(1-\gamma)K_0 e^{-E/kT_i}\Delta t_i, \quad \text{for} \quad \gamma \neq 1,$$

$$\ln n_i/n_{i-1} = -K_0 e^{-E/kT_i}\Delta t_i, \quad \text{for} \quad \gamma = 1. \quad (9)$$

By taking logarithms

$$\ln[(1/n_i^{\gamma-1}) - (1/n_{i-1}^{\gamma-1})] = \ln(\gamma-1)K_0\Delta t_i - (E/kT_i), \quad \text{for} \quad \gamma \neq 1,$$

$$\ln \ln[(n_{i-1})/n_i] = \ln K_0\Delta t_i - (E/kT_i), \quad \text{for} \quad \gamma = 1. \quad (10)$$

For constant Δt_i, the kinetic constants are derivable by plotting the left-hand side of Eqs. (10) *vs* $1/T_i$. A straight line with the proper activation energy will be obtained only for the correct γ, and this linearity criterion‡ can be used to derive the value of γ. Simple

‡ G. W. Brindley has proposed a similar relation recently in connection with annealing of cold work in copper. (Report of a Conference on Strength of Solids, University of Bristol, July 7–9, 1947; The Physical Society, London, 1948, p. 95). His equations are, however, incorrect, as he did not take into account the fact that for each Δt_i the relaxation starts at an n_{i-1} different from n_0.

calculations show, however, that it is essential to cover a wide temperature range accurately, because otherwise portions of the curved lines may be interpreted as straight, leading to incorrect values for E, γ, and K_0.

In practice, the main difficulty arises in obtaining sufficiently accurate data to be able to use the linearity criterion of Eqs. (10). The errors are cumulative in such an experiment, because each n_i depends on the n_{i-1} of the previous step. In general, Eqs. (10) are far more useful in predicting the response to such a heat treatment than in determining the fundamental kinetic constants. Obviously, such an analysis cannot be carried out if the $p=f(n)$ relation is not known.

In connection with tempering experiments, we may point out some of the difficulties encountered when a distribution in E is introduced to describe the process. According to Vand[5] a $\Delta n/\Delta T$ *vs* T plot, obtained from a tempering experiment, is an approximation to the shape of the distribution function in E. A plot of $\Delta n/\Delta T$ *vs* T is shown for a process of discrete $E(E/k=15{,}000°\text{K})$ in Fig. 7 for two heating rates. It is evident that a spread of about 10 percent at half-height is to be expected for a discrete activation energy because of the mathematical approximations introduced by Vand in evaluating a distribution function. From such an analysis one would conclude, therefore, the existence of a distribution in E when, in fact, the activation energy is discrete. If the question of a distribution in E *vs* discrete E arises, it is best settled on the basis of a phenomenological analysis of isothermal curves.

ACKNOWLEDGMENT

The authors are greatly indebted to Mr. J. M. Nuding and Mrs. Phyllis Stello for their assistance in designing the pulse-annealing apparatus and in carrying out the experimental measurements, and to Miss Ethel Stratton for performing many of the calculations required.

APPENDIX. RELATION BETWEEN DIFFERENTIAL ANNEALING CURVE MINIMA AND THE ACTIVATION ENERGY

The general rate Eq. (1) is no longer simply integrable when the experimental conditions are such that the temperature is a function of the time. Formally, we have

$$\int_{n_0}^{n}(dn/n^{\gamma})=-K_0\int_0^t e^{-E/kT}dt. \tag{A1}$$

If the temperature-time history can be approximated by a constant rate of heating, one can write

$$T=T_0+\alpha t \tag{A2}$$

and

$$dT=\alpha dt,$$

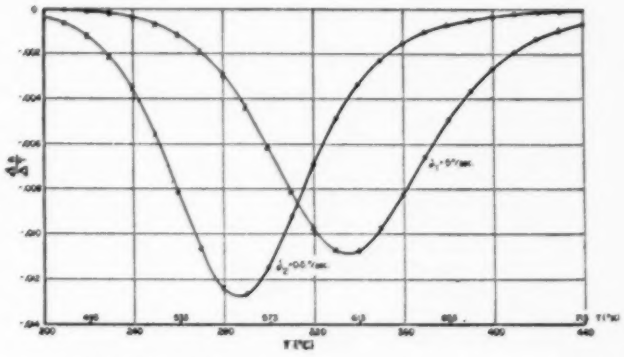


FIG. 7. Distribution function for E according to Vand's method.

where α is the temperature rise per unit time. Substitution in Eq. (A1) and integration gives

$$n_0^{1-\gamma}-n^{1-\gamma}=(1-\gamma)(K_0/\alpha)(E/k)\int_{kT_0/E}^{kT/E}e^{-1/T'}dT'. \tag{A3}$$

The differential curve measures $\Delta p/\Delta T$ as a function of T. First, we calculate dn/dt:

$$dn/dt=\alpha(dn/dT)=-K_0n^{\gamma}e^{-E/kT}. \tag{A4}$$

Substituting for n^{γ} from Eq. (A3) gives

$$dn/dt=-\left[n_0^{1-\gamma}-(1-\gamma)(K_0/\alpha)(E/k)\times\int_{kT_0/E}^{kT/E}e^{-1/T'}dT'\right]^{\gamma/(1-\gamma)}K_0e^{-E/kT}. \tag{A5}$$

The condition for maximum dn/dt is given by

$$(d/dT)(dn/dt)=0.$$

Using Eqs. (A5) and (A3) and denoting by T_m the temperature at which the maximum in dn/dt occurs gives, after considerable simplification, the relation

$$n^{1-\gamma}=(\gamma K_0/\alpha)e^{-E/kT_m}(k/E)T_m^2,$$

or

$$n_0^{1-\gamma}=(\gamma K_0/\alpha)e^{-E/kT_m}(kT_m^2/E)+(1-\gamma)(K_0/\alpha)(E/k)\int_{kT_0/E}^{kT_m/E}e^{-1/T'}dT' \tag{A6}$$

is obtained. The condition for maximum can be expressed, of course, in terms of p_0 by $p_0=f(n_0)$. However, for any given material and activation state n_0, and therefore p_0, is constant. Thus,

$$(1/\alpha)\left\{T_m^2e^{-E/kT_m}+[(1-\gamma)/\gamma](E/k)^2\int_{kT_0/E}^{kT_m/E}e^{-1/T'}dT'\right\}=\text{constant}. \tag{A7}$$

A change in the heating rate α produces a change in T_m. From this shift in T_m, the value of E may be approximately deduced as follows. For a first-order reaction the second term in Eq. (A7) is zero, and for higher orders the change in the second term is usually negligible relative to the first, over the range of E/T of interest. Approximately, then,

$$(1/\alpha)T_m^2e^{-E/kT_m}=\text{constant}, \tag{A8}$$

which immediately leads to Eq. (7) of the text.

Kinetics of Ordering in the Alloy AuCu*

G. J. DIENES
Atomic Energy Research Department, North American Aviation, Inc., Downey, California
(Received January 22, 1951)

The kinetics of ordering in AuCu has been studied using electrical resistivity as the index of order. A purely phenomenological analysis of isothermal resistivity-time curves showed that ordering in this alloy is a process characterized by a single constant activation energy of 29 kcal. For further kinetic analysis, it is necessary to transform the raw data into isothermal order parameter-time curves. This was accomplished via Muto's theoretical relation between resistivity and the Bragg-Williams order parameter, S. A consistent kinetic picture was obtained in terms of S, with the ordering process accurately described by a simple third-order rate equation with all the kinetic constants fully evaluated. The usual ln (rate constant) *versus* $1/T$ plot gave an activation energy of 28.5 kcal in excellent agreement with the phenomenological activation energy. Further experiments showed that the proposed kinetics predicts correctly the response to complex time temperature sequences.

I. INTRODUCTION

THE work described in this paper deals with some of the fundamental aspects of the rate of ordering in the alloy AuCu. The objective of the research program on copper-gold alloys in this laboratory is the quantitative comparison of reaction rates in specimens disordered by (a) quenching from high temperature, (b) cold work, and (c) irradiation with high energy particles. Such studies are of considerable importance in investigating the fundamental nature of radiation damage.† The ordering reaction in AuCu was chosen for this purpose, because prior work indicated that, at least in certain temperature ranges, the ordering process may be a single-step one of well-defined activation energy. Sufficient information could not be found in the literature available to the writer to permit a quantitative description of the ordering process, and, therefore, rather detailed kinetic studies were carried out in an attempt to fill this gap. The purpose of this paper is to describe and analyze these results, which are of general interest in the field of solid-state kinetics. Investigations of the effect of cold work and cyclotron irradiations upon ordering rates are now in progress and will be discussed in later papers.

The thermodynamics and statistics of order-disorder transformation have been investigated in considerable detail.[1] Far less work has been done on the kinetics of the ordering and disordering process. Sykes and Evans'[2] early work on $AuCu_3$, based on electrical resistivity measurements, indicates that the kinetics of transformation in this alloy is complex with the ordering taking place in at least two steps. This conclusion was confirmed by Siegel's study using the elastic modulus as an index of degree of order.[3]

In Gorsky's early work[4] on AuCu the ratio of the lattice parameters, a/c, was measured as the structure changed from face-centered cubic to tetragonal during ordering. The a/c *versus* time data appear to be rather rough in terms of accuracy and comparability at different temperatures because of lack of equivalence in starting points. From these data, by an analysis analogous to that given in Sec. III of the present paper, it appeared that ordering in AuCu may be essentially a single process with an activation temperature, E/k, in the neighborhood of 11,000°K. Further kinetic interpretation could not be made, since the connection between a/c and the degree of order is unknown and is almost certain to be quite complex.

More recently, Harker's[5] x-ray studies indicated a continuous growth process for the ordering of AuCu. Borelius,[6] Nyström,[7] and Borelius, Larsson, and Selberg[8] arrived at similar conclusions on the basis of their calorimetric, resistometric, and x-ray studies. In the 225° to 325°C temperature range the tetragonal phase apparently grows continuously from the disordered cubic phase. These workers obtained an activation temperature of about 18,000°K. It was thought that a more detailed kinetic study would be particularly profitable in this temperature range where the ordering in AuCu is apparently a single-step process.

II. EXPERIMENTAL

The kinetics of ordering was followed by resistivity changes. The sample, in the form of a thin ribbon, was quickly heated by an electric current to a chosen annealing temperature, held at this temperature for the desired length of time, and followed by rapid cooling

* This work was supported by the AEC under contract AT-11-1-GEN-8. Presented at the Los Angeles Meeting of the American Physical Society, December 28 through 30, 1950.

† For recent declassified articles on radiation damage in solids see F. Seitz, Disc. Faraday Soc., No. 5, 271 (1949); and J. C. Slater, J. Appl. Phys. **22**, 237 (1951).

[1] For reviews see F. C. Nix and W. Shockley, Revs. Modern Phys. **10**, 1 (1938); and R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, London, 1949), pp. 563–607. For recent work see J. M. Cowley, Phys. Rev. **77**, 669 (1950); J. Appl. Phys. **21**, 24 (1950); and Yin-Yuan Li, J. Chem. Phys. **17**, 447 (1949).

[2] C. Sykes and H. Evans, J. Inst. Metals **58**, 255 (1936).

[3] S. Siegel, J. Chem. Phys. **8**, 860 (1940).

[4] W. S. Gorsky, Physik. Z. Sowjetunion **6**, 69 (1934); **8**, 443 (1935).

[5] D. Harker, Trans. Am. Soc. Metals **32**, 210 (1944).

[6] G. Borelius, J. Inst. Metals **74**, 17 (1947).

[7] J. Nyström, Arkiv Fysik **2**, 151 (1950).

[8] Borelius, Larsson, and Selberg, Arkiv Fysik **2**, 161 (1950).

to a fixed temperature (25°C) for an in-place resistance measurement. This cycle, which gives an essentially square temperature pulse of any desired duration, was repeated until sufficient data were obtained for an isothermal annealing curve, or for any other suitable temperature-time history. The details of the apparatus, methods of resistivity measurement, and temperature measurement and control are fully described in another paper.[9] This technique, termed "pulse-annealing," has been found to be rapid, easy to carry out, and of good reproducibility in any temperature range where the heating and cooling times are a small fraction of the reaction time.

The polycrystalline AuCu samples were prepared as follows. The required weights of copper‡ and gold§ were melted in a graphite crucible in a hydrogen atmosphere. The sample was kept above 1100°C for about one hour with the container being shaken periodically, furnace cooled to a few hundred degrees C, and then air cooled to room temperature. The material was cold rolled into foils of about 0.005-inch thickness. The sample was then disordered by heating at 600°C for one hour followed by quenching. Chemical analysis of the sample gave 75.5 weight percent Au corresponding to 49.9 atomic percent Au.

Most of the experiments carried out were isothermal ordering studies consisting of the determination of resistance-time curves at several constant temperatures. Such isothermal curves are most amenable to theoretical analysis from the kinetic standpoint. Since it was highly desirable to use the same sample for many heat treatments, a standard technique of disordering was adopted consisting of a one-minute pulse at 600°C. This treatment brought the sample to a reproducible state as characterized by resistance measurement (R_0) at room temperature. This resistance did not correspond to that of a fully disordered sample but was about 5 percent lower corresponding to an initial resistivity of 13.5 micro-ohm cm (14.2 micro-ohm cm being the resistivity of a disordered sample of AuCu).

The results of the isothermal experiments are given as relative resistivity, $\rho/\rho_0(=R/R_0)$ *versus* t (t=time at reaction temperature) curves in Figs. 1 and 2, with all resistivity measurements taken at 25°C. The degree of reproducibility obtained is indicated by duplicate runs at 360°C. A more complex time sequence has also been studied, and a few long time furnace anneals carried out. These are discussed in Sec. V in connection with the kinetic analysis.

A fully quantitative analysis of the ρ/ρ_0 *versus* time curves is given in later sections. Qualitatively, Figs. 1 and 2 are interpreted as follows. From 250°C to 360°C essentially only ordering is taking place at a rate which increases with increasing temperature. The equilibrium

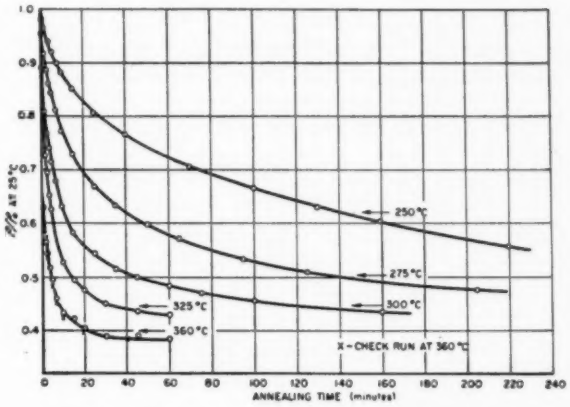


FIG. 1. Isothermal relative resistivity *vs* time curves for ordering in AuCu.

value at these temperatures after a long time anneal would correspond very nearly to complete order. As the reaction temperature is raised to 385°C and 390°C (Fig. 2), the apparent rate decreases and the final resistivity increases as compared to the results at 360°C. According to Nyström's work,[7] the increase in resistivity obtained at the higher temperatures is due to the presence of an orthorhombic ordered phase. Nyström's results also indicate that the rate of growth of the orthorhombic phase is slow relative to the tetragonal phase, and, consequently, this side reaction should lead to no complications below 360°C. The present paper is restricted to a study of the ordering kinetics in the

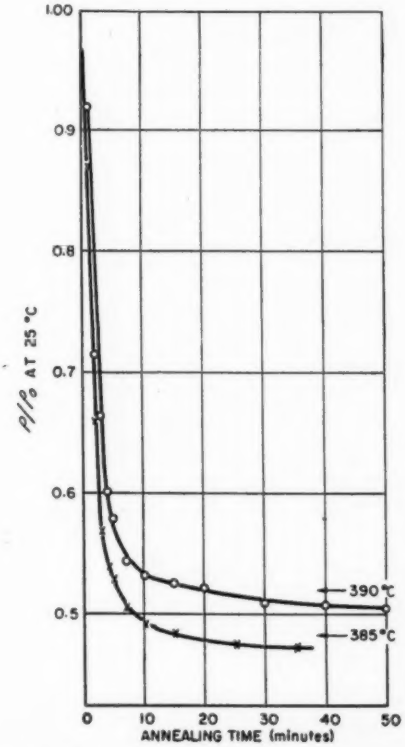


FIG. 2. Isothermal relative resistivity *vs* time curves for ordering in AuCu.

[9] Parkins, Dienes, and Brown, J. Appl. Phys. **22**, 1012 (1951), companion paper.

‡ Johnson, Mathey & Company, 99.999 percent pure copper.

§ Fine grade dental gold, S. S. White Company.

250°C to 360°C range where the influence of the growth of the orthorhombic form is negligible except perhaps at long times at the highest temperatures.

III. PHENOMENOLOGICAL ACTIVATION ENERGY

By a purely phenomenological approach, it is possible to determine the activation energy directly from the relative resistivity *versus* time data. For this purpose it is not necessary to know[9] the relation between the measured physical property (here ρ/ρ_0) and the fundamental quantity obeying a rate equation as long as a monotonic functional dependence may be assumed. As will be shown in Sec. IV, the kinetics of ordering can be formulated in a reasonable manner in terms of an order parameter, S. As far as the phenomenological activation energy is concerned, however, it is immaterial whether S is the true kinetic variable. It is only necessary to assume that a fundamental unit process for ordering exists and that this process, in terms of some suitable parameter, x, monotonically related to the measured physical property, obeys a rate equation. If the rate process is characterized by a single activation energy, the times and temperatures required to reach a given fixed fractional change in physical property must be related by the equation[9]

$$te^{-E/kT}=\text{constant} \tag{1}$$

regardless of the dependence of ρ on x. Conversely, conformation to Eq. (1) indicates that the rate process under investigation is a single-step process characterized by the activation energy, E.

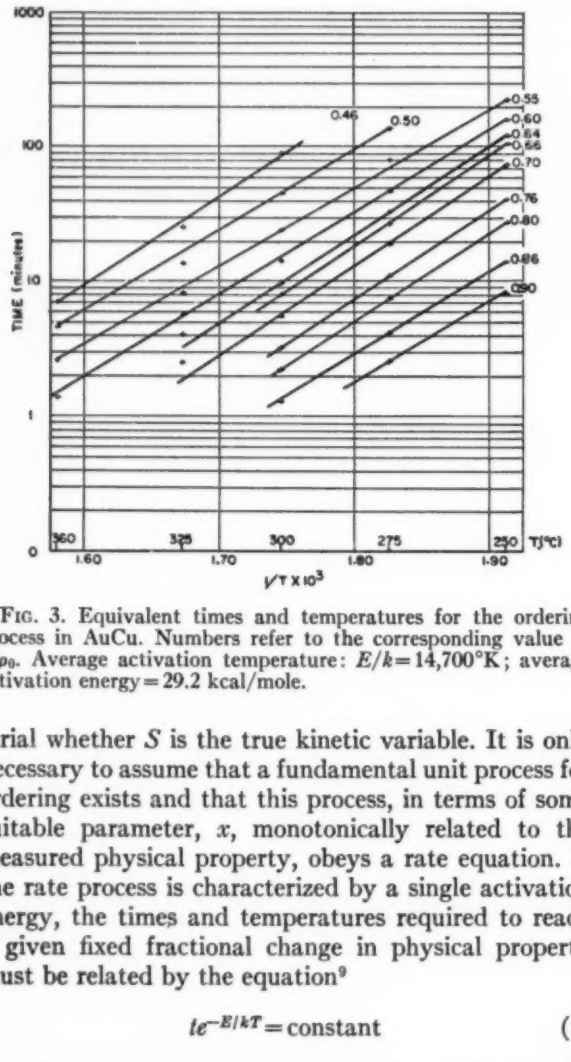


FIG. 3. Equivalent times and temperatures for the ordering process in AuCu. Numbers refer to the corresponding value of ρ/ρ_0. Average activation temperature: $E/k=14{,}700°\text{K}$; average activation energy = 29.2 kcal/mole.

In Fig. 3 the data are plotted as $\log_{10}t$ *versus* $1/T$ lines corresponding to various constant values of ρ/ρ_0 in accordance with Eq. (1). The lines are quite straight with slopes independent of the value of ρ/ρ_0 as required by the theory[9] underlying Eq. (1). There is some scatter in the slopes but no trend. From these lines an average activation temperature of 14,700°K was obtained.

The analysis so far indicates that one is justified in seeking a kinetic interpretation based on a single process of ordering characterized by a single activation energy. No further progress can be made, however, on a purely phenomenological basis. In order to determine rate constants and orders of reaction, it is necessary to know the connection between the physical property measured (ρ/ρ_0) and the degree of order, S, which, as will be shown, describes the kinetics of the process. In the following sections the transformation from ρ/ρ_0 to S, and the kinetics of ordering in terms of S will be discussed.

IV. RELATION BETWEEN RESISTIVITY AND DEGREE OF ORDER

Muto[10] investigated theoretically the resistivity of a superlattice alloy as a function of the degree of order. In this investigation, which is analogous to Nordheim's theory of the resistivity of alloys,[11] a quadratic relation was derived for resistivity as a function of the Bragg-Williams order parameter, S. Muto's final result may be written

$$\rho=\rho'+\rho''S+\rho'''S^2, \tag{2}$$

where ρ is the resistivity of the alloy. Two relations are immediately available for determining the adjustable constants in Eq. (2), namely,

$$\begin{aligned} &\text{at } S=0, \quad \rho=\rho_{\text{disordered}}, \\ &\text{at } S=1, \quad \rho=\rho_{\text{ordered}}. \end{aligned} \tag{3}$$

In order to evaluate all the constants, a third relation is assumed,

$$(d\rho/dS)_{S=0}=0. \tag{4}$$

There is good experimental justification for this last step, as ρ is known to be rather insensitive to changes in the degree of order near $S=0$. Combining Eqs. (2), (3), and (4) leads to the following relation, which no longer contains any adjustable constants,

$$(\rho-\rho_{\text{ord}})/(\rho_{\text{dis}}-\rho_{\text{ord}})=1-S^2, \tag{5}$$

with all the resistivities measured at the same temperature. It is perhaps of interest to point out that if $(1-S^2)$ is factored into $(1-S)(1+S)$, it is easy to see that changes in the $(1-S)$ factor account for most of the change in resistivity. Thus, Muto's refined scattering theory for resistivity is not very different from a simple scattering theory according to which one would assume

[10] R. Muto, Inst. Phys. Chem. Research Sci. Papers (Tokyo) **30**, 99 (1936).

[11] L. Nordheim, Ann Physik **9**, 607, 641 (1931); Also see F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), pp. 541–545.

a linear relation between changes in resistivity and the variation in the fraction of disordered atoms.

Most of the experimental support for relating resistivity to the Bragg-Williams long-range order comes from investigations of the $AuCu_3$ superlattice alloy.[12,13] The presence of differently oriented ordered domains is indicated also in AuCu by the calorimetric studies of Borelius and co-workers[8] as well as by the x-ray investigations of Harker[5] and of Hultgren and Tarnopol.[14] It seems likely, therefore, that also in this alloy the resistivity depends primarily on long-rather than short-range order. Estimates of average diffusion length during quench or during a very short anneal at reaction temperature point in the same direction, namely, that short-range order is essentially established before the kinetic measurements are started.‖

Further work is definitely needed in this field for a better understanding of the connection between physical properties and the lattice positions of the atoms. For example, simultaneous x-ray (lattice parameter as well as superlattice lines) and resistivity studies carried out on the same specimen of AuCu as a function of order would be extremely helpful.

V. KINETICS OF ORDERING IN AuCu

The results of the previous section enable one to transform relative resistivity-time curves into average degree of order-time curves which then can be subjected to kinetic analysis. Using $\rho_0=13.5$, $\rho_{dis}=14.2$, $\rho_{ord}=4.4$, the ρ/ρ_0 *versus* time curves of Fig. 1 were transformed into the S *versus* time curves of Fig. 4. As explained in Sec. II, the kinetic measurements start at $S_0=0.265$ corresponding to $\rho_0=13.5$. The results are not sensitive to the exact choice of ρ_0.

The kinetics of the ordering process is formulated as follows. In an AB superlattice alloy, the number of A and B atoms and the number of α- and β-sites are equal. Let r be the fraction of correctly occupied sites, i.e., the fraction of ordered atoms. The fraction of disordered atoms is given by $(1-r)$, with $\frac{1}{2}\leq r\leq 1$. The order parameter, S, is then given by $S=2r-1$, with $0\leq S\leq 1$. The phenomenological analysis of Sec. III has shown that, up to 360°C, the ordering takes place in a continuous manner with a constant activation energy. Application of reaction rate theory to this process suggests then that the rate of ordering is proportional to some power of the fraction of disordered atoms. Thus,

$$dr/dt=K(1-r)^{\gamma}, \tag{6}$$

and by substitution of S from $r=\frac{1}{2}S+\frac{1}{2}$, the kinetic relation becomes

$$dS/dt=K(1-S)^{\gamma}. \tag{7}$$

One has to determine now the order of reaction, γ, the rate constant, K, and the activation energy associated with K. The experimental data of Fig. 4 were found to be very well described by $\gamma=3$. For $\gamma=3$, integration yields

$$[1/(1-S)^2]-[1/(1-S_0)^2]=2Kt. \tag{8}$$

The experimental data are plotted according to Eq. (8) in Figs. 5 and 6. While there is some scatter around the straight lines, it is evident that a third-order kinetic equation describes the results very well. Deviations toward low values at high temperatures and long times are attributable to the growth of the orthorhombic phase.[7] The influence of the orthorhombic transformation is negligible over the main course of the reaction, and rate constants are reliably determined by the slopes of the straight lines of Figs. 5 and 6. With time expressed in minutes, the rate constants given in Table I were obtained.

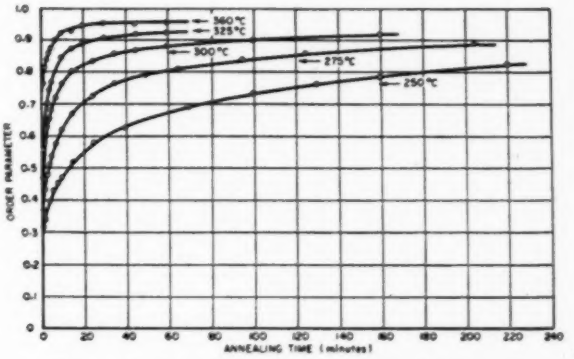


FIG. 4. Isothermal order parameter *vs* time curves for ordering in AuCu.

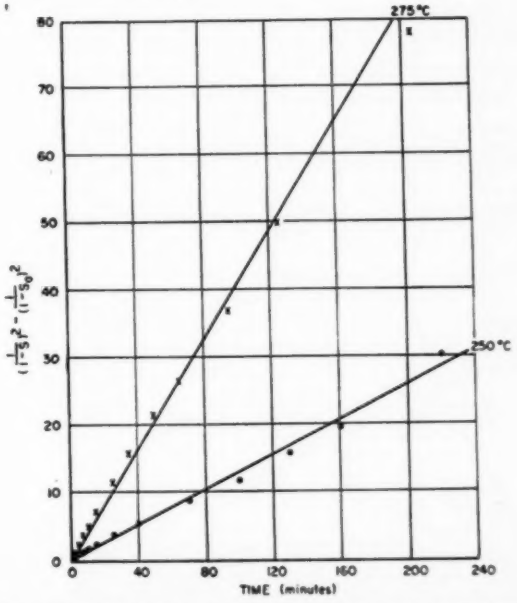


FIG. 5. Data of Fig. 4 plotted as a third-order reaction.

[12] C. Sykes and F. W. Jones, Proc. Roy. Soc. (London) **157A**, 213 (1936); **166A**, 376 (1938).
[13] C. Sykes and H. Evans, J. Inst. Metals **58**, 255 (1936).
[14] R. Hultgren and L. Tarnopol, Trans. Am. Inst. Mining Met. Engrs. **133**, 228 (1939).
‖ The writer is indebted to Dr. L. Guttman, who reviewed an earlier version of this paper, for clarification on these points as well as for calling the writer's attention to several important questions of interpretation.

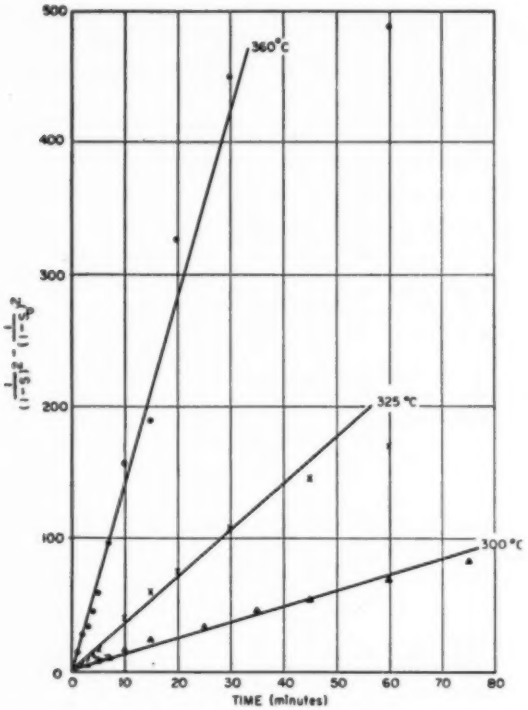


FIG. 6. Data of Fig. 4 plotted as a third-order reaction.

The kinetic data obtained at 250°C are plotted according to a second- and fourth-order reaction in Fig. 7. The nonlinearity of these curves, in comparison with the 250°C line of Fig. 6, shows very clearly the superiority of a third-order reaction, over those of other order, in describing the data. A similar comparison is valid at the other temperatures.

The logarithm of K is plotted against the reciprocal of the absolute temperature in Fig. 8. A good straight line can be drawn through these points, giving an activation temperature of 14,300°K. This in excellent agreement with the phenomenological activation energy determined in Sec. III from the raw relative resistivity data, as it, indeed, must be for consistency of analysis. Expressed in calories, the activation energy, or perhaps more precisely the heat of activation, for ordering in AuCu is 28.5 kcal per mole.

Expressed as a complete equation for K, the line of Fig. 10 gives

$$K=4.46\times10^{10}e^{-14300/T}. \tag{9}$$

The frequency factor of this equation appears to be a reasonable number. Further interpretation will have to wait for an elucidation of the atomic mechanism involved.

TABLE I. Rate constants for ordering of AuCu.

Rate constant, K (min^{-1})	Temperature, (°C)
0.065	250
0.204	275
0.61	300
1.78	325
7.50	360

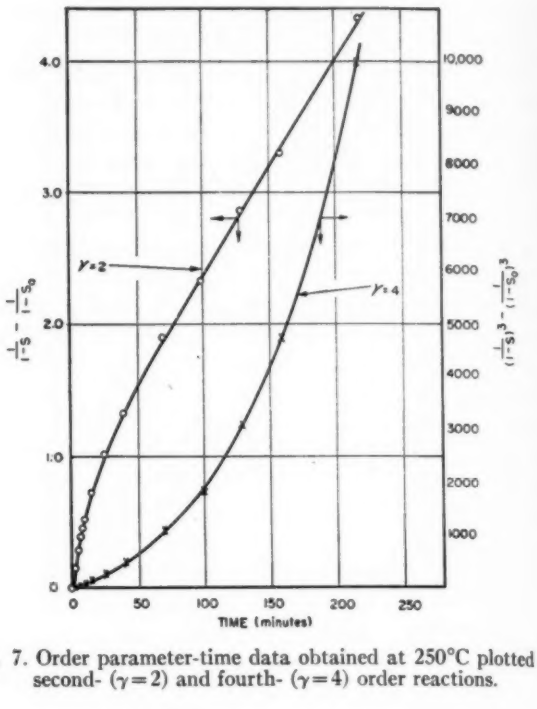


FIG. 7. Order parameter-time data obtained at 250°C plotted as second- ($\gamma=2$) and fourth- ($\gamma=4$) order reactions.

It is desirable to check a little further on the kinetics proposed here. Equation (9) was derived from data obtained on samples of an identical initial degree of order, S_0. Thus, one parameter in the kinetic equation (Eq. (9)) was kept constant, and it is important to know whether the proposed kinetics predicts the cor-

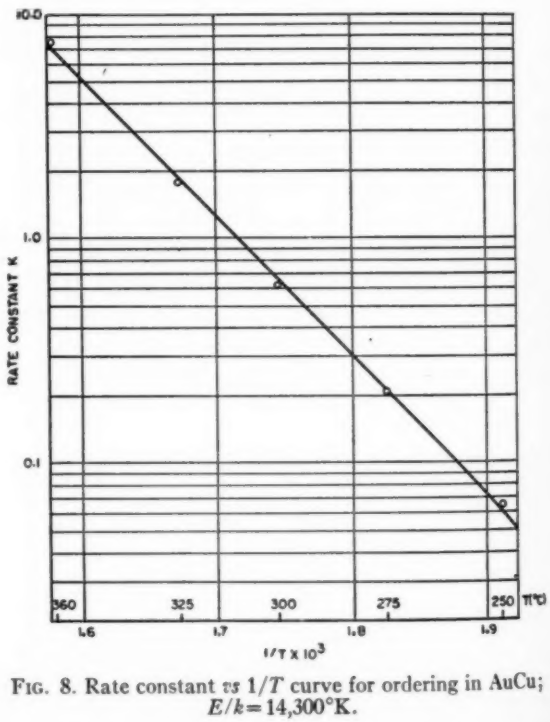


FIG. 8. Rate constant vs $1/T$ curve for ordering in AuCu; $E/k=14{,}300$°K.

rect behavior with respect to initial degree of order. To investigate this point, the following experiment was run by the technique described in Sec. II. The sample was allowed to order for 30 minutes at 250°C, with readings taken every 10 minutes. These points checked closely with the 250°C curve given in Fig. 1. At the end of this treatment, S was 0.597. The reaction temperature was changed to 275°C, and the ordering was followed for 70 minutes, giving a final S of 0.826. At this point, the temperature was changed to 300°C, and the reaction followed for 40 additional minutes. The data are shown in Fig. 9.

These data can be analyzed according to Eq. (9) using the proper S_0, namely, $S_0=0.597$, at 275°C, and $S_0=0.826$ at 300°C. The corresponding plots are shown in Fig. 10, with the rate constants obtained from these runs indicated on the graph. Excellent agreement is shown with previously determined K values (see Table I and Fig. 8). The proposed kinetics evidently predicts correctly the response to such a relatively complex heat treatment.

A few long-time furnace annealing experiments were also run. In these cases, the sample was initially essentially completely disordered, and, consequently, S_0 was taken to be zero in all calculations. From the kinetically determined rate constants, S was calculated for any given heat treatment and the results compared to the experimentally determined values as derived from resistivity measurements made at the end of the long-time anneals. This comparison is shown in Table II. Good agreement is observed, further substantiating the proposed kinetic picture. Since these samples were initially completely disordered, the data of Table II also indicate that the ordering reaction in AuCu is not characterized by an induction period measurable over the temperature range covered in these experiments.

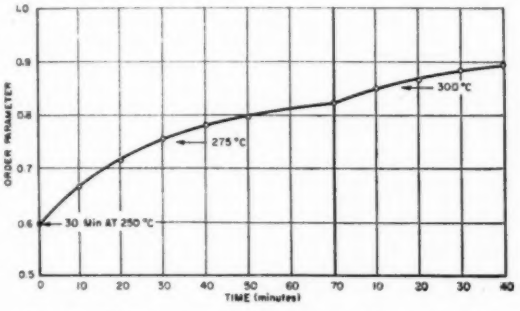


FIG. 9. Order parameter-time curves for a complex temperature-time history.

It is also of interest to compare the predictions of the kinetic equations with typical tempering experiments. By a tempering treatment the following temperature-time history is meant. The sample was subjected to temperature pulses of fixed duration, Δt_i, at successively higher temperatures, T_i. Two such experiments were performed using 5-minute pulses and varying the temperature in 10°C and 20°C steps, respectively. On the basis of Eqs. (8) and (9), the response of the material to such a heat treatment is easily calculable.¶ For any one of the pulses the following relation holds;

$$\log\{[1/(1-S_i)^2]-[1/(1-S_{i-1})^2]\} = \log 2K_0\Delta t_i-(E/2.3kT_i), \quad (10)$$

where S_i = degree of order at the end of the ith pulse, S_{i-1} = degree of order at the end of the $(i-1)$th pulse, Δt_i = duration of pulse, T_i = temperature of ith pulse.

The calculated and experimental points are compared, according to Eq. (10), in Fig. 11. Quite good agreement is observed, particularly if it is kept in mind that in such an experiment each step depends strongly on the result of each previous step and that, consequently, errors are to a large extent cumulative.

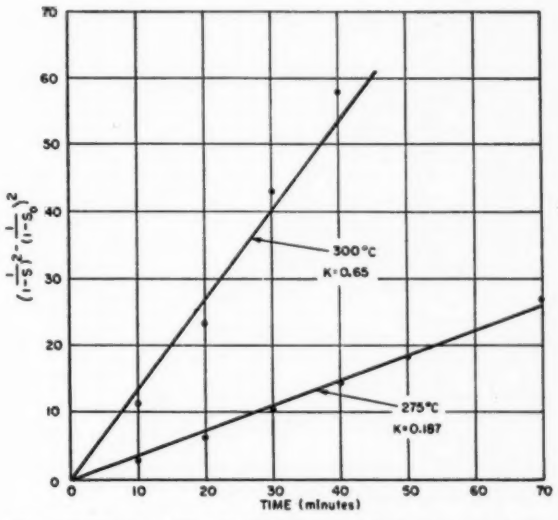


FIG. 10. Data of Fig. 9 plotted as a third-order reaction.

VI. DISCUSSION

The kinetic analysis presented in Sec. V is internally consistent as shown by the agreement of the activation energy derived from the kinetic study with that determined from the purely phenomenological analysis. Further, the proposed kinetics predicts correctly the response to complex time-temperature sequences. Consequently, the proposed kinetics is a useful, and at least

TABLE II. Long-time furnace anneals.

Heat treatment	S Calculated	S Experimental
20 hr at 250°C	0.920	0.930 0.945
100 hr at 250°C	0.964	0.970 0.985
20 hr at 300°C	0.975	0.973 0.977
100 hr at 300°C	0.989	0.988 0.983

¶ For a discussion of the interpretation of various types of annealing experiments see reference 9.

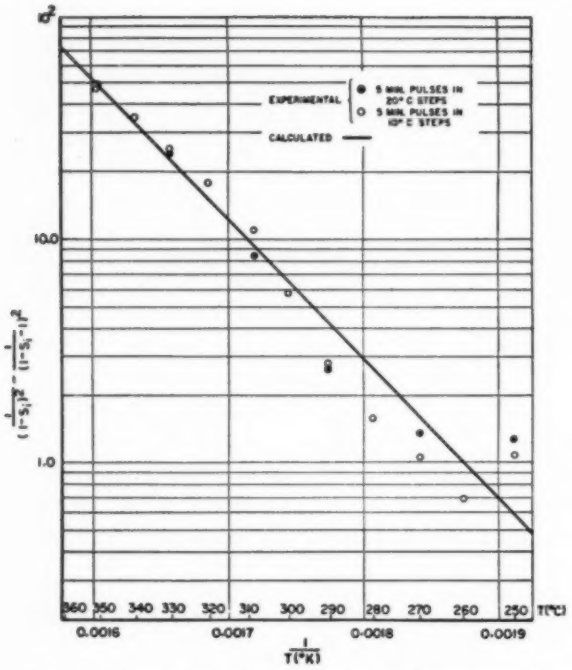


FIG. 11. Change in order as a function of temperature during tempering experiments.

phenomenologically, a correct description of the ordering process in AuCu.

The value of the activation energy obtained in this study is somewhat lower than the value given by Borelius.[8] It is felt, however, that the data of Fig. 3 offer strong support for the value derived from the experimental resistivity-time curves. It should be reemphasized that the activation energy, since it can be derived by a purely phenomenological analysis of the isothermal resistivity data, is entirely independent of the subsequent kinetic analysis. The magnitude of the activation energy indicates that diffusion phenomena play an important role in this order-disorder transformation. The activation energy for ordering, 28.5 kcal, is within experimental error identical to the activation energy for diffusion of Cu in AuCu alloys, namely 27.4 kcal, as given by Jost.[15]

According to the results presented, the ordering kinetics in AuCu is a simple one below 360°C and does not appear to proceed by a multistage nucleation and growth process. At least, no measurable induction period could be observed in these experiments. Further, the ordering process does not obey the type of time law expected on the basis of nucleation and growth as developed by Borelius and co-workers.[8] Rather, it has been shown that the data are accurately represented by a chemical rate equation of the third order. The physical significance of this result is not yet clear, and further theoretical and experimental work is necessary to elucidate the mechanism of this order-disorder transformation. In the interim, the results of the present investigation are fully usable as a correct phenomenological description of the ordering process in AuCu.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the stimulating discussions held with M. M. Mills, P. Duwez, and W. E. Parkins during the course of this work. The author is indebted to M. Tarpinian and Mrs. Phyllis Stello for carrying out the necessary experiments, and to Miss Ethel Stratton for performing the required calculations.

[15] W. Jost, Z. physik Chem. **21B**, 158 (1933).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

Flow of Gases in Porous Media

L. H. WILSON,* W. L. SIBBITT,† AND M. JAKOB‡
Purdue University, Lafayette, Indiana
(Received January 29, 1951)

This paper presents a brief theoretical review of the viscous and free molecular flow phenomena. The equations which describe these two types of flow are applied to flow through porous media and are combined to yield a semi-empirical equation which may be used not only for the viscous and free molecular flow, but also for the transition region between these. Experimentally determined limiting values of the ratio of molecular mean free path to pore diameter are reported. These ratios may be used to indicate the nature of flow in a porous media provided the pore size is known, or they may be used to determine pore size experimentally.

INTRODUCTION

WITH the recent developments in industrial use of low pressure filtering and the sweat-cooling application in rocket work using porous media, a knowledge of flow characteristics in porous media is required. The fundamental theoretical considerations necessary to characterize viscous and free molecular flow in capillary tubes have been known to physicists for a long time; however, the molecular flow is not so familiar to engineers. Therefore, this paper starts with a brief presentation of these different types of flow.

TYPES OF FLOW IN CAPILLARY SPACES

Rupert (1),§ in his paper on viscosity determination, states that flow through porous filters exhibits normal gas dynamical behavior and the filter behaves like an infinitely long capillary—thus, Poiseuille's equation is sufficient to describe the flow. This viewpoint needs considerable clarification, since the flow would not be of the gas-dynamical nature if the pore size were sufficiently small or the pressure sufficiently low. The present authors found, when attempting to determine high temperature viscosity data experimentally by allowing gas to flow through a porous graphite test section at atmospheric pressure, that the gas exhibited no viscosity effects within a temperature range from 70°F to about 3000°F. Obviously, the flow was of a free molecular nature rather than the normal viscous flow.

There are essentially four distinct types of fluid flow to be considered. They are (a) normal gas-dynamic flow, (b) viscous slip flow, (c) free molecular flow, and (d) the transition zone. Tsien (2) differentiates between these various regions by assuming limiting values of the ratio λ/δ, where λ is the molecular mean free path and δ is the boundary layer thickness. Although this type of differentiation is useful for aerodynamic purposes, it has no significance for general flow conditions. Each type flow will be considered separately.

* DuPont Research Fellow, Purdue University, Lafayette, Indiana.
† Associate Professor of Mechanical Engineering, Purdue University, Lafayette, Indiana.
‡ Research Professor of Mechanical Engineering, Illinois Institute of Technology, Consultant in Heat Research, Armour Research Foundation, Consultant in Heat Transfer, Purdue University, Lafayette, Indiana.
§ Numbers in parenthesis refer to bibliography at end of paper.

Gas Dynamic Flow

At normal pressures the volumetric rate of flow through cylindrical tubes can be expressed by Poisiuille's law:

$$\dot{V}=(\pi r^4/8L\mu)(p_1-p_2). \tag{1}$$

The assumptions in the derivation of the equation are that the velocity of flow at the wall is zero, the velocity profile is identical along the tube, and μ is a constant. In applying this equation to flow through porous media $\pi r^4/8L$ is a constant depending upon the geometry of the porous plate and must be determined experimentally. Here, r is a fictitious radius and L a fictitious length of a tube which would replace the porous plate regarding its flow resistance. From the kinetic theory explanation of the mechanism of flow, the resistance forces are entirely due to molecular momentum interchange and the viscosity is independent of pressure.

Viscous Slip Flow

It is observed that in flow through tubes at relatively low pressures there is an apparent decrease in the coefficient of viscosity. This effect is ascribed to a slipping of the gas at the walls of the containing vessel. If this velocity is taken as u_1 instead of zero, the resulting tangential force F on the wall is assumed to be proportional to u_1 and the surface area

$$F=Ku_1A, \tag{2}$$

where K is called the coefficient of external friction. Making a force balance on a differential length of the tube dL,

$$-\pi r^2dp=F=2\pi rKu_1dL, \tag{3}$$

one then obtains

$$u_1=(r/2K)dp/dL. \tag{4}$$

If this is used as the lower velocity limit in the derivation of Poiseuille's equation, the integration yields

$$\dot{V}=\pi r^4(p_1-p_2)(1+4\xi/r), \tag{5}$$

where

$$\xi=\mu/K=(\mu/p_m)(\tfrac{1}{2}\pi g_cRT)^{\frac{1}{2}}. \tag{6}$$

The symbol ξ is called the slip coefficient and represents the ratio of internal to external friction of the gas.

Free Molecular Flow

The physical concept of slip is applicable only in the case where the layer of gas flowing is many mean free paths thick so that ordinary viscous flow exists in the more distant parts. The situation is altogether changed where the gas layer is thin enough or the density low enough so that the collisions of the molecules with each other are extremely rare in comparison with the collisions of the molecules with the walls of the vessel. Assuming that the velocities of the molecules conform to the maxwell distribution law, it can be shown by a balance of momentum transfer and pressure drop in a tube of length L that

$$\dot{V}=(\pi/2)^{\frac{1}{2}}(r^3/L)(g_cRT)^{\frac{1}{2}}(p_1-p_2)1/p_m. \quad (7)$$

This type of flow is called free molecular flow. The volume of gas flow is independent of viscosity, but inversely proportional to pressure. Again, when applying this equation to a porous medium, $(\pi/2)^{\frac{1}{2}}r^3/L$ is a constant of geometry which must be determined experimentally.

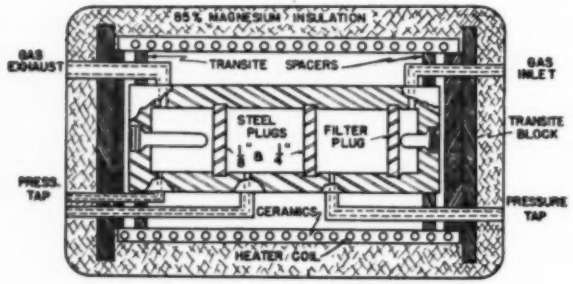


FIG. 1. Apparatus for fluid flow investigations.

The Transition Zone from Viscous to Free Molecular Flow

In the normal gas-dynamic flow the controlling factor is the transfer of momentum between molecules, while in the free molecular region the controlling factor is molecular bombardment of the walls of the vessel. There is obviously a transition zone between these types of flow where the collisions between molecules and the collisions of molecules with the wall are both of appreciable importance. There is no completely satisfactory theoretical presentation of the solution to this problem; however, Brown, Di Nardo, and associates (3) have presented a semi-empirical type of solution for flow through tubes, elbows, and valves.

For a rough approximation it may be stated that for flow through tubes, if $r/\lambda>100$, the normal viscous flow is predominate, while if $r/\lambda<1$, the free molecular flow is predominate.

SUMMARY OF THEORY

Equations (1), (5), and (7) describe the viscous, viscous slip, and free-molecular flow regions, respectively. It may be noted that when the value of the slip coefficient ξ in Eq. (5) is zero, the equation reduces to Eq. (1). It may also be noted that when the value of ξ is very large so that the 1 in Eq. (5) may be neglected, Eq. (5) reduces to Eq. (7). This observation implies that, in so far as mathematical description is concerned, the regions previously termed "viscous slip" and "transition zone" may be considered as one region, say, the "slip region," no matter what the physical mechanism in the two zones may be.

Equations (1) and (7) may be applied to porous media when the geometry factors have been determined by calibration; however, Eq. (5), which applies to the slip region, cannot be used directly, since the value of r in the additive term cannot be measured. This value can be determined from the calibrated constants in Eqs. (1) and (7). Equation 1 may be arranged in the form

$$\dot{V}=C_1(\Delta p/\mu), \quad (8)$$

where

$$C_1=\pi r^4/8L. \quad (9)$$

Equation (7) may be written in the form

$$\dot{V}=C_2(g_cRT)^{\frac{1}{2}}(p_1-p_2)1/p_m, \quad (10)$$

where

$$C_2=(\pi/2)^{\frac{1}{2}}(r^3/L). \quad (11)$$

Taking the ratio of C_1/C_2, one arrives at the following expression for r:

$$r=(C_1/C_2)(2)^{\frac{1}{2}}(\pi)^{\frac{1}{2}}. \quad (12)$$

From Eqs. (3), (11), and (12) the fictitious radius r and length L can be determined. Substituting this expression in Eq. (5) and simplifying, we obtain the following equation:

$$\dot{V}=(C_1/\mu)\Delta p[1+(C_2/C_1)(\mu/p_m)(g_cRT)^{\frac{1}{2}}], \quad (13)$$

or in terms of the slip coefficient

$$\dot{V}=(C_1/\mu)\Delta p(1+C_2\xi\bar{C}_1), \quad (14)$$

which is a quite general expression for the flow through a porous plate.

EXPERIMENTATION

There is obviously some critical value of the ratio of molecular mean free path to pore diameter which will determine the nature of flow in a porous medium. It is desirable to determine the critical ratio below which viscous flow will exist and the other critical ratio above which free molecular flow will exist.

In order to investigate the flow characteristics in a porous medium, a stainless steel apparatus as shown in Fig. 1 was constructed. The mean free path could be varied over a considerable range by controlling the pressure in the apparatus. Three stainless steel porous plugs were placed in series in the apparatus, the first of which acted as a filter. The second plug was $\frac{1}{4}$ inch thick, and the third $\frac{1}{8}$ inch thick. By subtracting the pressure drops across the latter two plugs, the end effects were

canceled. The porous plugs had a pore size of 4 to 6 microns, and a porosity of 10 to 30 percent. The gas permeability was given as 0.72 darcys.‖

For the low pressure runs the pressure drop was measured with a mercury manometer and the volumetric rate of flow was determined by measuring the discharge from a vacuum pump in a 4-liter calibrated burette. The pressure in the system was controlled by means of a cartesian manostat.

For the high pressure runs the pressure in the system was maintained at a constant value by throttling the gas when discharged from the system. The pressure drop was measured with a water manometer and the flow rate with a precision wet-test gas meter.

In this investigation all runs were made at room temperature. Helium and nitrogen were employed to give a check of the results.

RESULTS

Figures 2 and 3¶ show the results of the experimentation. Figure 2 shows the free molecular flow region for helium and nitrogen with $\Delta p/\dot{V}$ plotted against average pressure. The value of $\Delta p/\dot{V}$ is directly proportional to average pressure in the free molecular region and is represented by curve A for nitrogen and C for helium. Curves B and D represent the experimental values. The deviations of curves B and D from the free-molecular-flow curves indicate the beginning of the transition from free molecular to viscous flow. In the free molecular flow region the slopes of curves A and C should have a ratio of the square roots of the molecular weights of the gases under consideration. In this case the ratio should be 2.64; the measured value of the ratio was 2.50.

Figure 3 shows the viscous flow region for nitrogen and helium. Curves A and C represent viscous flow in which $\Delta p/\dot{V}$ is independent of pressure. Curves B and D are the experimental curves indicating the initial section of the transition zone from viscous to free molecular flow. The ratio of the ordinates of curves A and C should be equal to the viscosity ratio of the two gases at 75°F. These ratios were found to agree within 1 percent. From Fig. 2 one would expect that the molecular mean free path of the nitrogen at the point of tangency of curves A and B would be equal to the mean free path of the helium at the point of tangency of curves C and D. Actually, for helium the mean free path is between 5.4 and 6.5 microns and between 5.2 and 5.7 microns for nitrogen. Figure 3 indicates the transition from viscous flow to begin at a mean free path between 0.15 and 0.22 micron for nitrogen and between 0.25 and 0.30 micron for helium. These discrepancies were probably caused in part by experimental error and in part by the uncertainty of molecular

‖ A darcy is defined as 1 (cc/sec)/cm²/(atmos/cm).

¶ Since the pressure at points of deviation of experimental from theoretical curves in Figs. 2, 3, and 4 were the only points of interest, the units of measurement on the ordinates were chosen for convenience.

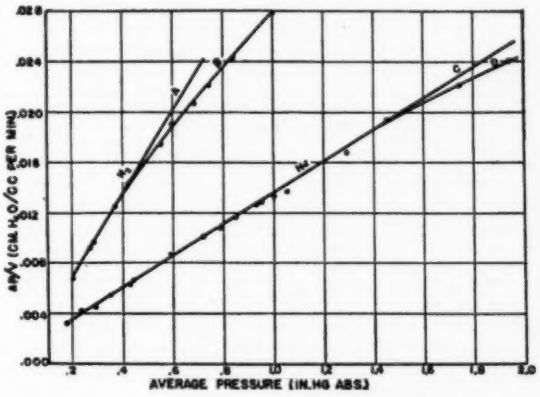


FIG. 2. The free molecular flow region.

mean free path data. These discrepancies were of little consequence for this investigation, however, since the pore size of the stainless steel was known only within an accuracy of ±20 percent.

Figure 4 shows a comparison of the experimental values with the curve given by Eq. (13). It may be noticed that the two curves agree within 5 percent except in the low pressure transition zone and the initial portion of the free molecular zone. The reason for this deviation is that the additive term in Eq. (5) is not of a magnitude such that the quantity 1 may be entirely neglected. The two curves would, of course, fit well in the lower pressure free molecular region.

The fictitious tube radius r was calculated from Eq. (12) using experimentally determined values of C_1 and C_2. The value of r thus calculated was found to be 3 microns, which agrees with the data given by the manufacturer of the porous plugs.

CONCLUSIONS

From the experimental data it may be concluded that viscous flow will exist in a porous medium if the ratio of the molecular mean free path to pore diameter is less than 0.025. This value was obtained using the minimum mean free path and assuming the pore diameter of the stainless steel to be 6 microns. Free molecular flow will exist if the ratio of the molecular mean free path to pore

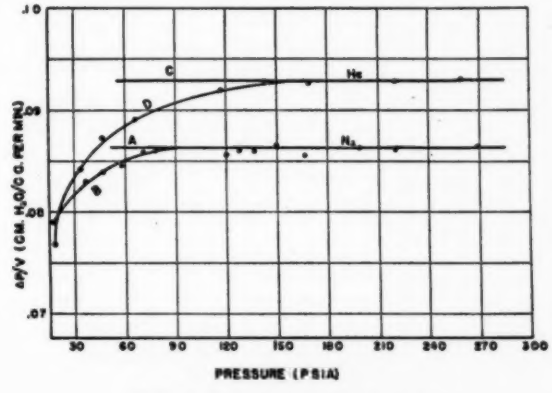


FIG. 3. The viscous flow region.

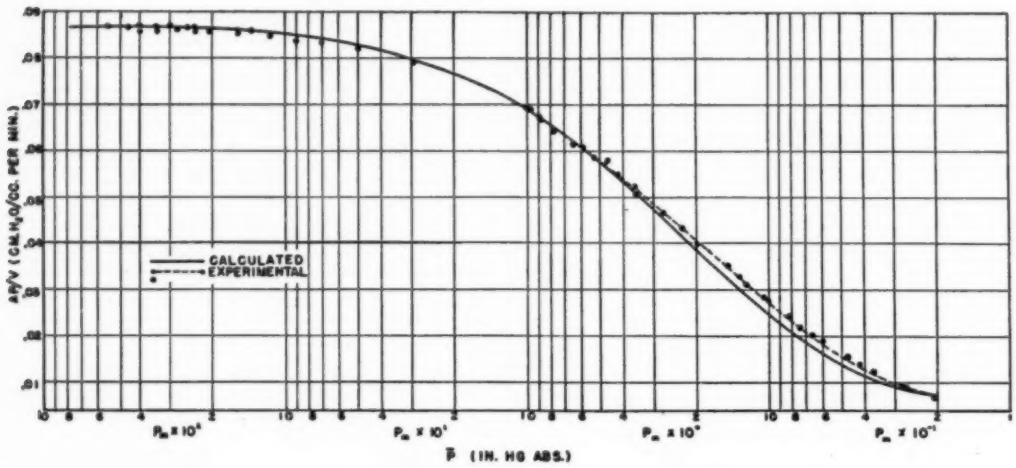


FIG. 4. Viscous, slip, and free molecular flow regions for nitrogen.

diameter is greater than 1.6. This value was obtained using the maximum mean free path and assuming the pore diameter is 4 microns.

Equation (12) may be used to determine a pore radius for fluid flow, the accuracy of the determination depending upon the instrumentation in the evaluation of C_1 and C_2. For engineering purposes Eq. (13) may be applied to the transition zone as well as the free molecular and viscous flow regions.

The principal features of the development presented here are that considerations of specular and diffuse molecular reflection as discussed in reference 3 need not be applied for flow in porous media. Also, the more accurate, but less direct, method proposed by Adzumi (7) may be avoided for the purpose of engineering calculations.

ACKNOWLEDGMENTS

The authors are indebted to the Du Pont Company for providing the funds which made this experimentation possible. Special thanks are also due Mr. Sidney Bernsen for his able assistance.

NOMENCLATURE

A	Area (ft²)
C_1	Geometrical constant (ft³)
C_2	Geometrical constant (ft²)
F	Force (lb)
g_c	Gravitational constant (ft/sec²)
K	Coefficient of external friction (lb sec/ft³)
L	Length (ft)
p	Pressure (lb/ft²)
p_m	$\frac{1}{2}(p_1+p_2)$ mean pressure (lb/ft²)
r	Tube radius (ft)
R	Individual gas constant (ft lb/lb °R)
T	Absolute temperature (°R)
u_1	Velocity (ft/sec)
$\dot{V}$	Volumetric rate of flow (ft³/sec)
δ	Boundary layer thickness (microns)
ξ	Slip coefficient (ft)
λ	Molecular mean free path (microns)
μ	Dynamic viscosity (lb sec/ft²)

BIBLIOGRAPHY

(1) Walter Ruppert, "The viscosity of gases," Messtech. **8**, 237-242 (1932).
(2) H. S. Tsien, "Superaerodynamics, mechanics of rarefied gases, J. Aeronaut. Sci. **13**, No. 12, 653 (1946).
(3) Brown, Di Nardo, Cheng, and Sherwood, "The flow of gases in pipes at low pressures," J. Appl. Phys. **17**, 802-813 (1946).
(4) S. Dushman, *Scientific Foundation of Vacuum Technique* (John Wiley and Sons, Inc., New York, 1949).
(5) E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938), first edition.
(6) L. B. Loeb, *The Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1934), second edition.
(7) H. Adzumi, "Flow of gases through porous diaphrams," Bull. Chem. Soc. Japan 12, 304 (1937).

Studies in Newtonian Flow. I. The Dependence of the Viscosity of Liquids on Temperature

ARTHUR K. DOOLITTLE
Carbide and Carbon Chemicals Company, A Division of Union Carbide and Carbon Corporation, South Charleston, West Virginia

(Received January 25, 1951)

It is shown that the Walther and Andrade equations, while very satisfactory for moderate ranges of temperature, are not adequate to define the dependence of the viscosity of liquid normal paraffins on temperature over extended ranges. One method of developing a viscosity-temperature relationship is presented that gives equations defining this dependence with satisfactory validity. The purpose of this paper is to show that the precise dependence of the viscosity of liquids on temperature is complicated. This is in contrast to the relatively simple relation for the precise dependence of the viscosity of liquids on free-space which will be presented in subsequent papers.

THE dependence of the viscosity of liquids on temperature has been studied by many investigators over a long period of years, and a large number of relationships involving these variables has been accumulated. A resumé of a number of these expressions relating viscosity to temperature may be found in a chapter by the author in Alexander's *Colloid Chemistry*.[1] None of these relationships has proven satisfactory over extended ranges of the variables.

In an effort to gain further insight into the nature of the flow process and of the dependence of viscosity on various factors, a study has been in progress in this laboratory for a number of years. Although solutions of resinous materials and homologous series of polar liquids (solvents) were at first studied, it became apparent as the work progressed that it would be necessary to deal with homologous series of compounds of a type that would be as free as possible from complications due to association. Accordingly, a series of ten very pure normal alkanes was prepared, ranging in molecular weight from 72 to 900; and the densities and viscosities were measured with great accuracy over a range of temperatures from −10°C to 300°C. The densities were measured in specially designed pycnometers; the viscosities in a carefully calibrated Ubbelohde viscometer. A description of the synthesis and measurement of these compounds, with an estimate of the accuracy of each measurement, is published elsewhere.[2] For convenience in making the subsequent calculations, the measured densities and viscosities were adjusted to their corresponding values at even temperatures, applying methods of interpolation described in the previous paper. The temperatures chosen were −10°, 20°, 50°, 100°, 150°, 200°, 250°, and 300°C. Table I lists the values of the densities and viscosities adjusted to these even temperatures, and Fig. 1 is a plot of the natural logarithms of these "observed" viscosity values against reciprocal absolute temperature.

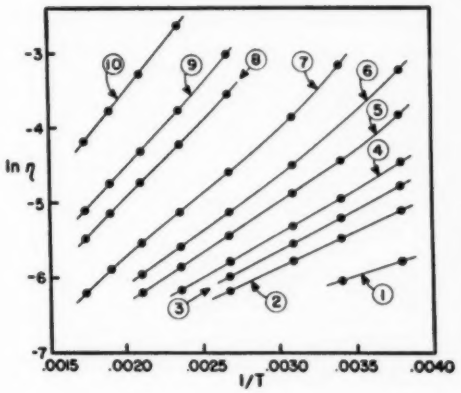
Many of the expressions proposed by investigators for relating the viscosity of liquids to temperature were examined critically by us in the light of this body of extensive and accurate data. Of all the expressions examined, the Walther equation,[3]

$$\log\log(\nu+0.8) = -n\log(T/T_1)+\log\log(\nu_{T_1}+0.8) = A\log T+B, \quad (1)$$

where ν = kinematic viscosity, T = absolute temperature, and the de Guzman-Andrade equation,[4,5]

$$\eta = Ae^{B/T} \text{ or } \ln\eta = B/T+\ln A, \quad (2)$$

where η = viscosity and T = absolute temperature, best reproduce the data over extended ranges of temperature. The Walther equation was considerably better at the low temperature end of the range for each compound studied. Reference to Table II will show the closeness of fit of these two equations in the case of



		M			M
①	C_5H_{12}	72	⑥	$C_{13}H_{28}$	184
②	C_7H_{16}	100	⑦	$C_{17}H_{36}$	240
③	C_8H_{18}	114	⑧	$C_{28}H_{58}$	395
④	C_9H_{20}	128	⑨	$C_{36}H_{74}$	507
⑤	$C_{11}H_{24}$	156	⑩	$C_{64}H_{130}$	900

FIG. 1. Logarithms of observed viscosities of normal alkanes *vs* reciprocal absolute temperature.

[1] J. Alexander, *Colloid Chemistry* (Reinhold Publishing Corporation, New York, 1950), Vol. VII, pp. 150–154.
[2] A. K. Doolittle and R. H. Peterson, J. Am. Chem. Soc. **73**, 2145 (1951).
[3] C. Walther, Proc. World Petroleum Congr., London **2**, 419 (1933).
[4] J. de Guzman, Anales soc. españ. fís. y quím. **11**, 353 (1913).
[5] E. N. da C. Andrade, Nature **125**, 309 (1930).

TABLE I. *n*-Paraffin hydrocarbons. Observed densities and viscosities adjusted to even temperature values.

m →			72.15			100.20			114.22			128.25			156.30		
t	T	1/T	ρ	η	lnη	ρ	η	lnη	ρ	η	lnη	ρ	η	lnη	ρ	η	lnη
−10	263.2	0.003800	0.6548	0.003114	−5.772	0.7087	0.006008	−5.115	0.7265	0.008314	−4.790	0.7408	0.01150	−4.465	0.7623	0.02167	−3.832
20	293.2	0.003411	0.6263	0.002404	−6.031	0.6839	0.004180	−5.477	0.7026	0.005479	−5.207	0.7177	0.007165	−4.939	0.7402	0.01186	−4.435
50	323.2	0.003094				0.6582	0.003100	−5.776	0.6782	0.003928	−5.540	0.6942	0.004968	−5.305	0.7180	0.007609	−4.878
100	373.2	0.002680				0.6119	0.002099	−6.166	0.6350	0.002554	−5.970	0.6532	0.003087	−5.781	0.6800	0.004384	−5.430
150	423.2	0.002364										0.6086	0.002136	−6.149	0.6393	0.002894	−5.845
200	473.2	0.002113													0.5947	0.002048	−6.191

m →			184.35			240.46			394.74			506.95			899.68		
t	T	1/T	ρ	η	lnη	ρ	η	lnη	ρ	η	lnη	ρ	η	lnη	ρ	η	lnη
−10	263.2	0.003800	0.7778[a]	0.03922[b]	−3.239	0.7780[a]	0.04209[a]	−3.168									
20	293.2	0.003411	0.7563	0.01883	−3.972	0.7575	0.02170	−3.830									
50	323.2	0.003094	0.7349	0.01118	−4.494	0.7232	0.01018	−4.587	0.7555	0.02872	−3.550	0.7667	0.04884	−3.019	0.7878[d]	0.1785[c]	−1.723
100	373.2	0.002680	0.6985	0.006001	−5.116	0.6880	0.005984	−5.119	0.7235	0.01460	−4.227	0.7357	0.02315	−3.766	0.7579	0.07144	−2.639
150	423.2	0.002364	0.6604	0.003765	−5.582	0.6512	0.003938	−5.537	0.6910	0.008856	−4.727	0.7048	0.01344	−4.309	0.7283	0.03757	−3.282
200	473.2	0.002113	0.6201	0.002615	−5.947	0.6119	0.002785	−5.884	0.6578	0.005920	−5.129	0.6730	0.008749	−4.739	0.6981	0.02292	−3.776
250	523.2	0.001911				0.5687	0.002030	−6.200	0.6226	0.004173	−5.479	0.6399	0.006083	−5.102	0.6680	0.01525	−4.183
300	573.2	0.001745															

[a] By extrapolation from −5.4°C. [d] By extrapolation from 109.89°C.
[b] By extrapolation from −5.48°C. [e] By extrapolation from 110.08°C.
[c] By extrapolation from 22.00°C.
Note. Viscosity, η, given in poises; density in g/ml (in vacuum).

n-heptadecane. Since both of these functions are transcendental, their constants were determined from our data by approximation methods. The "best lines" (or curves) were established as the ones for which the sums of the squares of the deviations of the viscosities (not their logarithms) gave minimum values.

The Walther equation, a double exponential type, is well known as the basis of the ASTM cross section paper for rectifying viscosity-temperature data of lubricating oils. It is obviously without theoretical significance, however, since it becomes meaningless when the kinematic viscosity, ν, approaches the value 0.2. The Andrade equation, on the other hand, is quite widely accepted as a fundamental relationship, and several theoretical or quasi-theoretical treatments have been advanced leading to this form of equation.

In spite of the good fit afforded by the Walther and the Andrade equations over moderate ranges of temperature, it is clear by reference to Table II that a corresponding degree of validity is not obtained when the temperature range under review is extensive. Deviations of four to twenty or more times the known accuracy of the measurements occur over practically the entire range of the measurements (280 centigrade degrees) when either of these equations is applied to our *n*-heptadecane data.

Since none of the relatively simple relationships for expressing the dependence of the viscosity of liquids on temperature that have been proposed in the literature are adequate to define the values as accurately as they can be measured, and for several other reasons, we were led to the belief that perhaps temperature is not a primary variable upon which the viscosity of liquids depends. It seemed more probable that viscosity varies in a regular manner with some property of liquids that is itself affected, in a more obscure way, by changes in temperature. An extensive study was therefore carried out in this laboratory to relate viscosity with the relative free-space, or more accurately, with the fractional increase in volume due to the thermal expansion of the liquid. This study will be reported in subsequent publications.

Before abandoning altogether the thought that temperature might be a direct variable, a considerable study was made of our normal paraffin data in an attempt to find a relationship, empirical or other, that would reproduce the measured values with satisfactory validity. The only purpose in presenting the present paper is to demonstrate (1) that even the best of the relatively simple viscosity temperature functions is inadequate to reproduce the measured values accurately over extended ranges of temperature and (2) that relationships that do reproduce the measured values with satisfactory validity are necessarily complicated and not likely to hold promise of fundamental significance.

In the process of developing a more accurate viscosity temperature relationship, a great many methods

TABLE II. Viscosity of *n*-heptadecane. Comparison of observed values with those calculated by Andrade[a] and Walther[b] equations.

			Andrade[a]		Walther[b]		Equation (4)	
T	η_{obs}	Percent[c] accuracy	η_{calc}	Percent error	η_{calc}	Percent error	η_{calc}	Percent error
293.2	0.04209	0.2	0.03948	−6.2	0.04297	+2.1	0.04030	−4.3
323.2	0.02170	0.2	0.02244	+3.4	0.02170	0.0	0.02165	−0.2
373.2	0.01018	0.2	0.01071	+5.2	0.009901	−2.7	0.01027	+0.9
423.2	0.005984	0.3	0.006084	+1.7	0.005754	−3.8	0.005978	−0.1
473.2	0.003938	0.5	0.003896	−1.1	0.003843	−2.4	0.003915	−0.6
523.2	0.002785	0.6	0.002717	−2.4	0.002801	+0.6	0.002762	−0.8
573.2	0.002030	0.8	0.002018	−0.6	0.002159	+6.4	0.002047	+0.8
Σ (Percent error)²→				87.3		73.3		20.9

[a] $\ln\eta = 1785(1/T) - 9.320$.
[b] $\log\log(\nu+0.8) = -3.6000\log T + 8.7855$.
[c] The assignment of accuracy values is described in reference 2.
Coefficients in the Andrade and Walther equations approximate least-square values as determined by repeated trial and error calculations.

of plotting the data were examined. None of the methods employed by others that were investigated by us appeared to offer much hope for success. Attempts to express the constant B in the Andrade equation as a function of temperature were particularly frustrating, as may be appreciated by reference to Fig. 2 from Table III. Only when the graphically determined slopes of the $\ln\eta$ vs $1/T$ curves were plotted against $\ln\eta$ was any regularity obtained. This method of plotting (see Fig. 3) gave a straight line (with slope 500) for all molecular weights at 100°C ($T = 373$°K). For most of the values applying to molecular weights 240 and lower, the points fell on straight lines (for each temperature) parallel to the above line. For all values applying to molecular weights above 240, the points fell more or less on straight lines that converged to a common point on the $T = 373$ line. It is not claimed that the parallel and converging lines represented in Fig. 3 accurately define these slopes. The relationship is obviously only approximate. The justification for using this method of developing a viscosity-temperature differential equation is simply that it represents the greatest regularity found after extensive study and that the integrated forms derived from this equation do reproduce the measured values quite satisfactorily.

From the geometry of Fig. 3, it can readily be shown that

$$d\ln\eta/d(1/T) = 500\ln\eta + 4023 + A(T-373). \quad (3)$$

For the parallel lines, $A = 4.54$; for the converging lines, $A = (500/T + 127)\cdot(-2\ln\eta - 5)$. Strictly speaking, the values that lie on the converging lines are those above the normal, $d\ln\eta/d(1/T) = -500\ln\eta - 750$, that is, the values of $d\ln\eta/d(1/T)$ when

$$500\ln\eta + 4023 + 4.54(T-373) > -500\ln\eta - 750,$$

or when

$$-\ln\eta > (4.54T/1000) + 3.08, \text{ lie on the converging lines.}$$

Actually, this definition puts one value of $m = 156$, three values of $m = 184$, and three values of $m = 240$ on the converging lines instead of on the parallel lines. The additional validity that might be gained by utilizing both values of A in the case of these three compounds does not justify the use of two equations

TABLE III. *n*-Paraffin hydrocarbons. Values of $d(\ln\eta)/d(1/T)$ determined graphically from $\ln\eta_{obs}$ vs $1/T$.

m→		72.15		100.20		114.22		128.25		156.30		184.35	
T	$1/T$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$
263.2	0.003800	−5.772	660	−5.115	932	−4.790	1093	−4.465	1260	−3.832	1650	−3.239[a]	2095
293.2	0.003411	−6.031	660	−5.477	938	−5.207	1055	−4.939	1180	−4.435	1443	−3.972	1741
323.2	0.003094			−5.776	940	−5.540	1044	−5.305	1150	−4.878	1362	−4.494	1563
373.2	0.002680			−6.166	940	−5.970	1040	−5.781	1145	−5.430	1315	−5.116	1482
423.2	0.002364							−6.149	1178	−5.845	1329	−5.582	1457
473.2	0.002113									−6.191	1417	−5.947	1505

m→		240.46		394.74		506.95		899.68	
T	$1/T$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$	$\ln\eta$	$d(\ln\eta)/d(1/T)$
293.2	0.003411	−3.168[a]	2279						
323.2	0.003094	−3.830	1931						
373.2	0.002680	−4.587	1727	−3.550	2239	−3.019	2486		
423.2	0.002364	−5.119	1679	−4.227	2065	−3.766	2252	−2.639	2714
473.2	0.002113	−5.537	1693	−4.727	1990	−4.309	2135	−3.282	2480
523.2	0.001911	−5.884	1765	−5.129	2052	−4.739	2155	−3.776	2438
573.2	0.001745	−6.200	1960	−5.479	2132	−5.102	2214	−4.183	2430

[a] By extrapolation.

TABLE IV. Comparison of observed viscosities with those calculated by

$$\ln\eta_{calc_1} = -4.54e^{500/T}[-\mathrm{Ei}(-500/T)] - 4.66 + C_1e^{500/T},$$
$$\ln\eta_{calc_2} = -3.31 - 25z(1+30z+30\cdot29z^2+\cdots+30!z^{30}) + C_2(27.1z)^{31}e^{1/z}, \text{ where } z = (1/27.1)(1+127/T).$$

$m\rightarrow$	72.15				100.20				114.22				128.25				156.30			
T	η_{obs}	η_{calc_1}	Δ_1	Percent acc[a]	η_{obs}	η_{calc_1}	Δ_1	Percent acc	η_{obs}	η_{calc_1}	Δ_1	Percent acc	η_{obs}	η_{calc_1}	Δ_1	Percent acc	η_{obs}	η_{calc_1}	Δ_1	Percent acc
263.2	0.003114	0.003106	−0.3	0.4	0.006008	0.005961	−0.8	0.2	0.008314	0.008329	+0.2	0.2	0.01150	0.01156	+0.5	0.2	0.02167	0.02140	−1.2	0.2
293.2	0.002404	0.002421	+0.7	0.4	0.004180	0.004142	−0.9	0.2	0.005479	0.005453	−0.5	0.2	0.007165	0.007137	−0.4	0.1	0.01186	0.01186	0.0	0.1
323.2					0.003100	0.003103	+0.1	0.4	0.003928	0.003926	−0.1	0.3	0.004968	0.004939	−0.6	0.3	0.007609	0.007619	+0.1	0.2
373.2					0.002099	0.002118	+0.9	0.5	0.002554	0.002563	+0.4	0.5	0.003087	0.003090	+0.1	0.4	0.004384	0.004393	+0.2	0.3
423.2													0.002136	0.002145	+0.4	0.5	0.002894	0.002897	+0.1	0.5
473.2																	0.002048	0.002065	+0.8	0.6

$m\rightarrow$	184.35				240.46				394.74				506.95				899.68			
T	η_{obs}	η_{calc_1}	Δ_1	Percent acc ...[b]	η_{obs}	η_{calc_1}	Δ_1	Percent acc	η_{obs}	η_{calc_1}	Δ_2	Percent acc	η_{obs}	η_{calc_2}	Δ_2	Percent acc	η_{obs}	η_{calc_2}	Δ_2	Percent acc
263.2	0.03922	0.03703	−5.6	0.1	0.04209	0.04030	−4.3	0.2												
293.2	0.01883	0.01862	−1.1	0.2	0.02170	0.02165	−0.2	0.2	0.02872	0.02907	+1.2	1.4	0.04884	0.04944	+1.2	1.5	0.1785	0.1814	+1.6	...[b]
323.2	0.01118	0.01120	+0.2	0.2	0.01018	0.01027	+0.9	0.2	0.01460	0.01471	+0.8	1.3	0.02315	0.02328	+0.6	1.4	0.07144	0.07158	+0.2	1.5
373.2	0.006001	0.006009	+0.1	0.4	0.005984	0.005978	−0.1	0.3	0.008853	0.008853	0.0	1.2	0.01344	0.01345	+0.1	1.4	0.03757	0.03740	−0.5	1.6
423.2	0.003765	0.003784	+0.5	0.4	0.003938	0.003915	−0.6	0.5	0.005920	0.005864	−0.9	1.3	0.008749	0.008687	−0.7	1.4	0.02292	0.02273	−0.8	1.5
473.2	0.002615	0.002615	0.0	0.6	0.002785	0.002762	−0.8	0.6	0.004173	0.004149	−0.6	1.5	0.006083	0.006042	−0.7	1.5	0.01525	0.01522	−0.2	1.5
523.2					0.002030	0.002047	+0.8	0.8												
573.2																				

[a] The assignment of accuracy values is described in reference 2.
[b] Extrapolation too great to permit assignment of accuracy value.

to define the viscosity-temperature relationship for a single substance.

Integration of Eq. (3) for a constant value of A requires a table of logarithmic integrals, but is otherwise straightforward. The integrated form is

$$\ln\eta_{calc_1} = -Ae^{500/T}[-\mathrm{Ei}(-500/T)] - 4.66 + C_1e^{500/T}, \quad (4)$$

where $-\mathrm{Ei}(-500/T)$ is the logarithmic integral and C_1 is the molecular weight dependent constant of integration. The average values of C_1 were calculated to be as follows:

m	C_1
72	0.0885
100	0.1860
114	0.2361
128	0.2850
156	0.3773
184	0.4593
240	0.5996.

Integration of Eq. (3) with the variable value of A is complicated by the fact that the integral involves a series of 31 terms. Thus, if $z = 1/27.1(1+127/T)$, the integrated form becomes

$$\ln\eta_{calc_2} = -3.31 - 25(1+30z+30\cdot29z^2+30\cdot29\cdot28z^3 + \cdots+30!z^{30}) + C_2(27.1z)^{31}e^{1/z}. \quad (5)$$

The constant of integration, C_2, is again molecular weight dependent and the following average values were calculated from the data:

m	$C_2\cdot10^{12}$
395	3.858
507	4.641
900	6.558.

Values of viscosity calculated according to Eq. (4) for molecular weights up to 240, and values calculated according to Eq. (5) for molecular weights greater than 240 are as close to the observed values as the accuracy of the measurements in 65 percent of the cases (see Table IV). The difference between the calculated and the observed values exceeds the accuracy of the measurements by >1 percent in only 2 of the 49 cases. (Far better agreement would obviously have been obtained in these 2 cases if both Eqs. (4) and (5) had been used to cover the range instead of Eq. (4) alone.) It is evident, therefore, that our empirical equations do define the viscosity-temperature relationship of the n-alkanes with a high degree of validity. Comparison of Eq. (4) with the Andrade and with the Walther equations is given for n-heptadecane ($m=240$) in Table II.

SUMMARY

1. A differential viscosity-temperature relationship for the n-alkanes is developed. The coefficient of the temperature term in this differential equation is substantially constant for n-alkanes of molecular weights

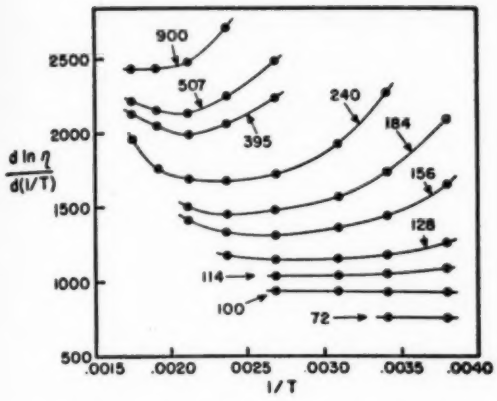


FIG. 2. Slopes of the $\ln\eta$ *vs* $1/T$ "lines" *vs* reciprocal absolute temperature.

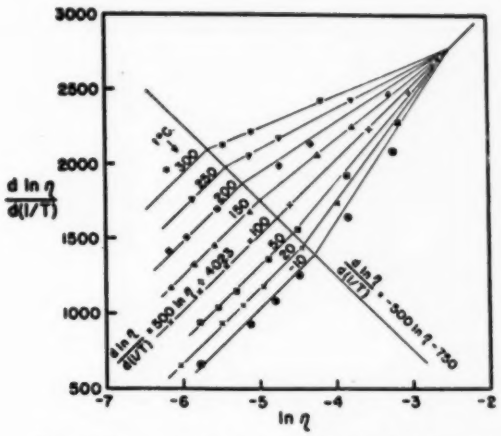


FIG. 3. Slopes of the $\ln\eta$ *vs* $1/T$ "lines" *vs* $\ln\eta$.

$\leqq 240$, but has a variable value for the higher molecular weight paraffins. Integration gives two equations, one applicable to molecular weights $\leqq 240$, the other to molecular weights >240.

2. The deviations from the measured viscosities of values calculated according to these integrated forms are within the accuracy of the measurements in a majority of the cases. In only 2 of our 49 cases do these deviations exceed the accuracy of the measurements by >1 percent.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

On Creep and Relaxation. III

B. GROSS
The National Institute of Technology, Rio de Janeiro, Brazil

AND

H. PELZER
The Electrical Research Association, Greenford, Middlesex, England

(Received January 31, 1951)

The previous theory of linear reversible creep is extended and permanent (plastic) deformations are taken into consideration. General relations are established between the distribution functions of relaxation times of stress, and retardation times of strain. A particular type of distribution function is considered in detail and it is discussed in a qualitative way how the existence of a plastic component in creep affects the distribution function of relaxation times.

IN two previous papers[1] the general theory of linear relaxation was developed. In particular, the role of the distribution functions of relaxation times of stress and retardation times of strain was considered, and relations between the creep and the relaxation curves were obtained. The theory as exposed in these papers was confined to the case where all deformations were completely reversible. Plastic deformations and permanent set were excluded from the considerations. Neither have they been treated, so far as we know, in theoretical papers by other authors. This obviously represents a severe limitation. It is the purpose of the present paper to generalize the theory so as to include such phenomena. To keep in line with the former treatment, however, effects are still considered to be linear. This means that the plastic deformation produced by application of a constant load (stress) is supposed to increase with time at constant rate, and at any time to be strictly proportional to the load. This again represents a limitation for the theory, but is the only case amenable to a simple treatment.

[1] B. Gross, J. Appl. Phys. **18**, 212 (1947); **19**, 257 (1948)—in this paper referred to as I and II. See S. Whitehead, J. Sci. Instr. **21**, 73 (1944).

I. REVERSIBLE CREEP AND PLASTIC DEFORMATION

In the general case now under consideration, the deformation which a viscoelastic body experiences under application of a constant load (stress) contains three components:

(a) The instantaneous elastic deformation given by

$$P_0F/E,$$

where P_0 is the applied load, E the instantaneous modulus of elasticity, and F a form factor;

(b) The delayed reversible deformation,

$$P_0\psi(t)F/E,$$

where $\psi(t)$ is the creep function;

(c) The plastic irreversible deformation,

$$P_0\eta Ft.$$

The total deformation under constant load therefore can be written as:

$$y(t)=(F/E)P_0(1+\psi(t)+kt). \tag{1}$$

And $k=E\eta$.

The total deformation under varying load $P(t)$ follows from (1) as

$$y(t)=\frac{F}{E}\left(P(t)+\int_0^t P(\tau)\varphi(t-\tau)d\tau+\int_0^t kP(\tau)d\tau\right). \tag{2}$$

$\varphi(t)$ is the rate of creep given as $\varphi(t)=d\psi/dt$. The two integrals can be written as a single integral with the introduction of a new function

$$\varphi^*(t)=\varphi(t)+k.$$

Employing φ^*, the expression for the total deformation takes the same form as that given in I, Eq. (2). Physically the difference lies in the fact that now the rate of creep contains a constant term, while formerly it went to 0 when time goes to ∞.

II. RELAXATION

Introduction of a component of plastic deformation does not change the formal expression for stress relaxation under constant strain. Therefore this relation is written in the same form as before (I, Eq. (4)).

$$P(t)=(E/F)y_0[1-\bar{\psi}(t)], \tag{3}$$

where $\bar{\psi}(t)$ is the relaxation function. For varying stress this becomes

$$P(t)=\frac{E}{F}\left(y(t)-\int_0^t y(\tau)\bar{\varphi}(t-\tau)d\tau\right), \tag{4}$$

where $\bar{\varphi}(t)=d\bar{\psi}(t)/dt$ is the rate of relaxation.

III. APPLICATION OF THE LAPLACE TRANSFORM

The coexistence of Eqs. (2) and (4) implies a relation between the functions φ and $\bar{\varphi}$. This can again be obtained with the aid of the laplace transform. Multiply both (2) and (4) with $\exp(-pt)$ and integrate over t from 0 to ∞. Observe that

$$\int_0^\infty e^{-pt}dt\int_0^t P(\tau)\varphi(t-\tau)d\tau = \left(\int_0^\infty e^{-pt}P(t)dt\right)\left(\int_0^\infty e^{-pt}\varphi(t)dt\right), \tag{5}$$

a similar relation holding for the integral involving y and $\bar{\varphi}$.[2] Introducing again the laplace transforms of φ and $\bar{\varphi}$ by

$$L(p)=\int_0^\infty e^{-pt}\varphi(t)dt \tag{6a}$$

$$\bar{L}(p)=\int_0^\infty e^{-pt}\bar{\varphi}(t)dt, \tag{6b}$$

there results easily

$$\bar{L}(p)=[L(p)+k/p]/[1+k/p+L(p)]. \tag{7}$$

Equation (7) already permits a useful conclusion. It follows from it that when $p=0$, $\bar{L}(0)=1$. But from (6b), $\bar{L}(0)=\bar{\psi}(\infty)$. Therefore

$$\bar{\psi}(\infty)=1, \tag{8}$$

whatever the form of $\psi(t)$. Then, according to (4a), under constant deformation the stress will always decay to 0: A substance suffering plastic deformation will always show complete relaxation.

It has been shown in I that mathematical and physical considerations justify the representation of both φ and $\bar{\varphi}$ as laplace transforms,

$$\varphi(t)=\int_0^\infty e^{-pt}\rho(s)ds \tag{9a}$$

$$\bar{\varphi}(t)=\int_0^\infty e^{-pt}\bar{\rho}(s)ds. \tag{9b}$$

Substituting Eq. (9) in Eq. (6) gives again

$$L(p)=\int_0^\infty \frac{\rho(s)ds}{s+p} \tag{10a}$$

$$\bar{L}(p)=\int_0^\infty \frac{\bar{\rho}(s)ds}{s+p}. \tag{10b}$$

Rotating the argument through $(-\pi)$ gives

$$L(pe^{-i\pi})=\int_0^\infty \frac{\rho(s)ds}{s-p}+i\pi\rho(p), \tag{11a}$$

$$\bar{L}(pe^{-i\pi})=\int_0^\infty \frac{\bar{\rho}(s)ds}{s-p}+i\pi\bar{\rho}(p). \tag{11b}$$

Exactly the same transformations have been employed already in *I*. A discussion of Eq. (11) has been given elsewhere.[3] Substitution of (11) into (7) gives finally,

$$\bar{J}+i\pi\bar{\rho}=[J-k/p+i\pi\rho]/[1-k/p+J+i\pi\rho], \tag{12}$$

where

$$J=\int_0^\infty \frac{\rho(s)ds}{s-p}, \tag{13}$$

[2] R. V. Churchill, *Modern Operational Methods in Engineering* (McGraw-Hill Book Company, Inc., New York, 1944), p. 37.

[3] B. Gross, Phil. Mag. Ser. 7 **51**, 543, (1950).

an analogous equation holding for $\bar{J}$ and $\bar{\rho}$. Separation into real and imaginary component gives

$$\bar{\rho}(p)dp=\rho(p)dp/[(1-k/p+J)^2+\pi^2\rho^2]. \quad (14)$$

The integrals (11) and (13) are principal values.

IV. THE DISTRIBUTION FUNCTIONS

The interpretation of creep and relaxation experiments in terms of distribution functions has become widely accepted.[4] These are introduced by the equations,

$$\psi(t)=\beta\int_0^\infty F(\tau)(1-e^{-t/\tau})d\tau, \quad (15)$$

$$\bar{\psi}(t)=\bar{\beta}\int_0^\infty \bar{F}(\tau)(1-e^{-t/\tau})d\tau. \quad (16)$$

$F(\tau)$ is the distribution function of retardation times of strain, and $\bar{F}(\tau)$ that of relaxation times of stress. To be distribution functions they must satisfy the normalization condition,

$$\int_0^\infty F(\tau)d\tau=1; \quad \int_0^\infty \bar{F}(\tau)d\tau=1. \quad (17)$$

Comparison of (15), (16), and (9) gives the relation between the auxiliary functions ρ and the distribution functions F.

$$\beta F(\tau)d\tau=-[\rho(p)/p]dp, \quad p=1/\tau, \quad (18)$$

an analogous equation holding for $\bar{F}$. Substitution of (18) into (14) gives the relation between F and $\bar{F}$. Observing that from (8), (16), and (17),

$$\bar{\beta}=1, \quad (19)$$

there results

$$\bar{F}(\tau)d\tau=\frac{\beta F(\tau)d\tau}{\left[1-k\tau+\tau\cdot\int_0^\infty \frac{\beta F(u)du}{\tau-u}\right]^2+[\pi\beta\tau F(\tau)]^2}. \quad (20)$$

This is the fundamental equation of the present theory. It shows how $\bar{F}$ can be calculated when F is known. In consequence of the particular form of (20), this calculation is possible when F is given in analytical or numerical form. Possibly the integrations have to be carried out by approximate or numerical processes, but in principle they are always possible. The importance of (20) derives from the fact that a corresponding equation does not hold for the relation between the creep and the relaxation functions. If one of these functions is given, the other has to be calculated in a much more involved way than that given above.

[4] C. Zener, *Elasticity and Anelasticity of Metals* (Chicago University Press, Chicago, 1948).

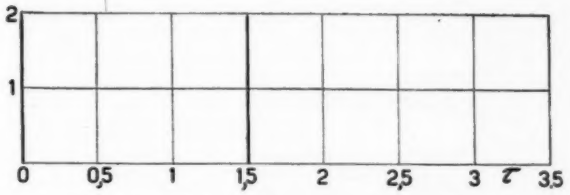


FIG. 1. Relaxation spectrum for $m=0$.

V. THE COLE-DISTRIBUTION

It is well known that a complete analogy exists between the anelastic properties of solids and the phenomena of dielectric absorption and dispersion in dielectrics.[5] For dielectrics, Cole and Cole[6] have shown that most data can be satisfactorily represented by a special function which is related to a transcendental known as the Mittag-Leffler function. The use of this function for the interpretation of relaxation data has already been proposed several times,[7] and Nolle[5] has found that it indeed yields an approximate representation of the results of measurements made with rubberlike materials, although in this case it apparently does not have the same general validity as for dielectrics. However it certainly is sufficiently flexible to permit at least a qualitative discussion of the influence exercised by the presence of plastic deformation. Therefore, in the following use will be made of it to develop the relation between the time distributions and to put into evidence more directly, the scope and possibilities of the present theory.

The ratio of mechanical displacement to mechanical force is analogous to the ratio of the electrical displacement to electromotive force (S. Whitehead, reference 1). Electric current is proportional to the derivative of electrical displacement. The Cole function for the absorption current therefore is analogous to a Cole function for the rate of creep. And the distribution function which Cole and Cole have derived from their data is identical with the distribution function of retardation times of strain $F(\tau)d\tau$. Thus we put, according to reference 6 (1941), Eq. (13),

$$F(\tau)d\tau=[\sin\pi m/(\tau^m+\tau^{-m}+2\cos\pi m)](d\tau/\pi\tau). \quad (21)$$

Here m is a parameter which expresses the width of the distribution, and

$$0<m<1. \quad (22)$$

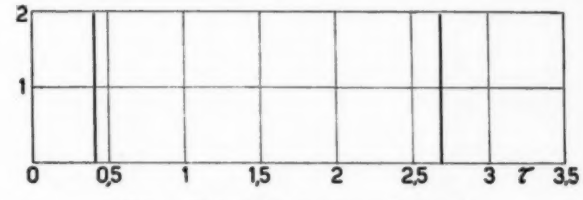


FIG. 2. Relaxation spectrum for $m=1$.

[5] A. W. Nolle, J. Polymer Sci. **5**, 1 (1950).

[6] K. S. Cole and R. H. Cole, J. Chem. Phys. **9**, 341 (1941); **10**, 98 (1942).

[7] B. Gross, J. Appl. Phys. **18**, 212 (1947). R. Sips, J. Polymer Sci. **5**, 55 (1950).

τ is actually a reduced time, but this obviously does not imply a restriction. Substitution of (21) into (20) and evaluation of the integral by means of the formulas (11) finally yields

$$\bar{F}(\tau)d\tau=\frac{1}{\pi}\frac{\pi\beta\tau F(\tau)}{[1-k\tau+\beta[\tau^m+\cos\pi m]/[\tau^m+\tau^{-m}+2\cos\pi m]]^2+[\pi\beta\tau F(\tau)]^2}\frac{d\tau}{\tau}. \tag{23}$$

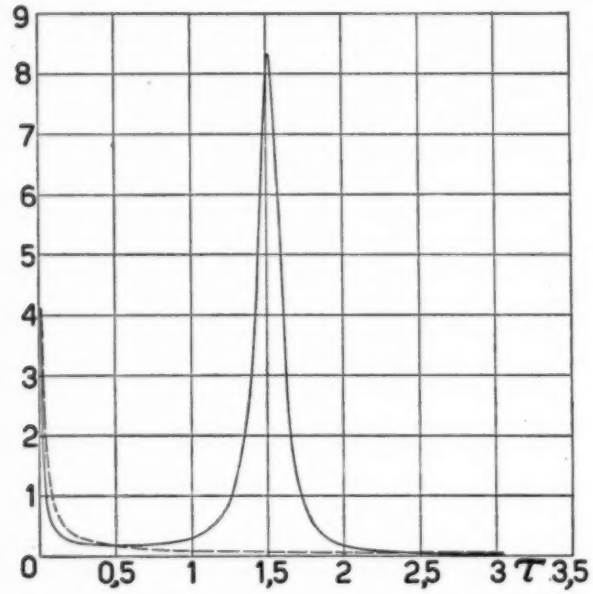


FIG. 3. Distribution functions $\pi\bar{F}(\tau)$ (solid curve) and $\pi F(\tau)$ (dashed curve) for $m=0, 1$.

For the discussion of this formula we consider different values of m.

(a) Wide Distribution

For $m\to 0$, F is approximately a $1/\tau$ distribution,

$$F(\tau)d\tau=md\tau/4\tau. \tag{24a}$$

Equation (24a) is valid only within a certain interval of τ, which depends on m. Outside this interval, that is

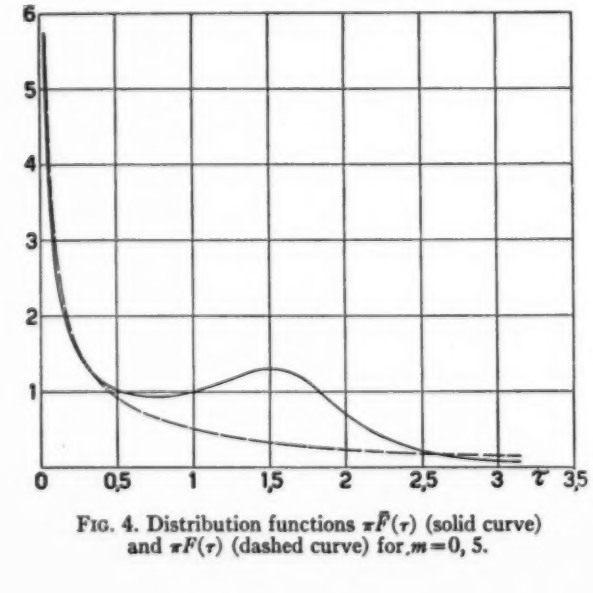


FIG. 4. Distribution functions $\pi\bar{F}(\tau)$ (solid curve) and $\pi F(\tau)$ (dashed curve) for $m=0, 5$.

for very small and very great values of τ, the factor multiplying $d\tau/\tau$ tends to 0 as it must, for otherwise the integral would diverge. The length of the interval increases with decreasing m, while the amplitude of F decreases. Then

$$\bar{F}(\tau)d\tau=\frac{\pi\beta m/4}{[k(\tau-\tau_1)]^2+[\pi\beta m/4]^2}\frac{d\tau}{\pi\tau}. \tag{24b}$$

For small values of τ this is essentially a $1/\tau$ distribution like (24a); for high values of τ, $\bar{F}$ goes to 0 with $1/\tau^3$.

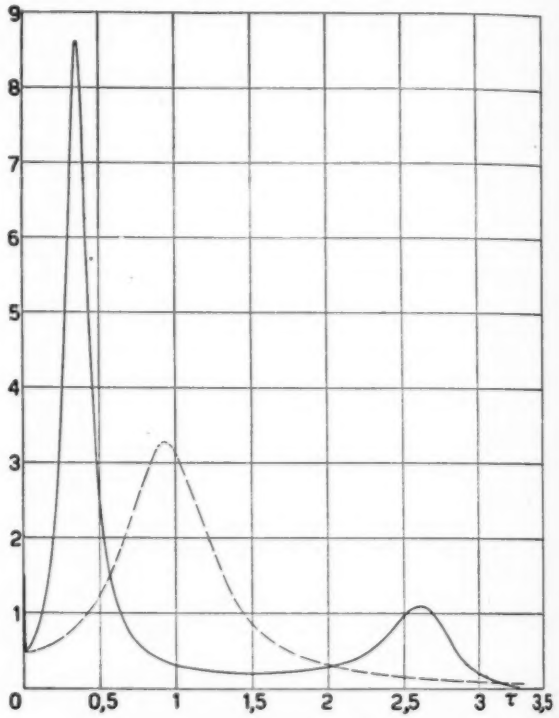


FIG. 5. Distribution functions $\pi\bar{F}(\tau)$ (solid curve) and $\pi F(\tau)$ (dashed curve) for $m=0, 9$.

Superposed there appears a peak which in the limit $m=0$ becomes a line with center at

$$\tau_1=(1+\beta/2)/k. \tag{25}$$

This follows from the form of the factor of $d\tau/\tau$ in (24b) which is that of a δ-function,[8]

$$\delta(\tau-\tau_1)=\lim_{\epsilon\to 0}\frac{1}{\pi}\frac{\epsilon}{(\tau-\tau_1)^2+\epsilon^2}. \tag{26}$$

[8] For this representation and other properties of the δ-function see B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge University Press, London, 1950), p. 71.

The $\bar{F}$-spectrum is shown in Fig. 1. When considering this figure one must observe that only in consequence of the linear scale the $1/\tau$ distribution prevailing at small values of τ appears like a line at $\tau=0$. In the logarithmic scale which would be proper for this region $\bar{F}$ would increase linearly with decreasing τ over a wide interval.

(b) Sharp Distribution

For $(m-1)\to 0$. F is approximately given by

$$F(\tau)d\tau=\frac{\pi(1-m)}{[(\tau-1)^2/\tau]+\pi^2(1-m)^2}\frac{d\tau}{\pi\tau}. \quad (27a)$$

There is $F(0)=0$. Then comparison with (26) shows that in the limit $m=1$ this is a sharp line with center at

$$\tau=1. \quad (28)$$

The corresponding $\bar{F}$ is given by

$$\bar{F}(\tau)d\tau=\frac{\beta\pi(1-m)}{[k(\tau-\tau_1)(\tau-\tau_2)]^2+[\beta\pi(1-m)]^2}\frac{d\tau}{\pi\tau}. \quad (27b)$$

Again one has $\bar{F}(0)=0$. τ_1 and τ_2 are the roots of the equation,

$$\tau_{1,2}=\frac{(1+\beta+k)\pm[(1+\beta+k)^2-4k]^{\frac{1}{2}}}{2k}. \quad (29)$$

In the limit $m=1$ the expression (27b) corresponds to that of a δ-function with argument $(\tau-\tau_1)(\tau-\tau_2)$, representing two lines with centers at τ_1 and τ_2 (Fig. 2).

(c) Medium Wide Distributions

In Figs. 3 to 5 are shown curves for $m=0, 1$; $m=0, 5$; $m=0, 9$. For the sake of simplicity the calculations were made with the values $\beta=1$ and $k=1$, as above.

These curves show that the existence of plastic deformation always produces a separate peak in the relaxation spectrum. This peak is very sharp when the retardation spectrum is sharp too, and in this case the relaxation spectrum has indeed two peaks. For a medium wide retardation spectrum the peak is rather blunt but still clearly discernible. However when the spectrum of retardation times becomes extremely wide and is approaching the $1/\tau$ distribution, the peak in the relaxation spectrum becomes very pronounced again.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

An Electrostatically Induced Permanent Memory

CHARLES F. PULVARI
The George Washington University, Washington, D. C.
(Received February 5, 1951)

The possibility of using a ferroelectric substance (barium titanate) as a medium for storing information is described. Preliminary experiments were performed to determine the necessary conditions for inscribing and erasing electrostatic preorientation in barium titanate ceramics at a temperature somewhat below the Curie point. The results suggest the possibility of a memory device that can be operated in the range of audio-frequencies or higher.

INTRODUCTION

WHAT we call "memory" is a link between time and space. Events occurring in time have to be transformed into a space arrangement or pattern corresponding to the given time function. If this transformation is stable, it becomes a permanent memory.

The storing and recovering of information is a typical communication problem with some unique features added.

A memory device has three characteristic operations: writing, reading, and erasing. The main aspects on which one particular device is chosen are speed of operation, cost per stored information unit (binary digit), space requirement per information unit, and reliability, or the reciprocal of the frequency of errors. For example, if space is at a premium, one must consider data such as that given in Table I, which presents a comparison among a few of the important memory devices.

TABLE I. The specific capacity of some important memory devices.

Type of memory	Approximate specific capacity
Human brain	6.7×10^6 units/cm^3
Microfilm	3.5×10^5
Dictionary	3.5×10^4
Magnetic tape	1.8×10^4

The specific capacity of a memory may be defined as the number of information units per unit volume.

Most of the electrostatic methods are of the dynamic and fast working type, in which a regeneration process permits the information to be retained.[1–5]

[1] A. V. Haeff, Electronics **20**, 80–93 (1947).
[2] J. A. Rajchman, "Selectron," *Mathematical Tables and Other Aids to Computation* (October, 1947), Vol. 2, pp. 359–361.
[3] Eckert, Luhoff, and Smohar, Proc. Inst. Radio Engrs. **38**, 498–510 (1950).
[4] R. C. Hergenrother and B. C. Gardner, Proc. Inst. Radio Engrs. **38**, 740–747 (1950).
[5] Jensen, Smith, Mesner, and Flory, RCA Review, **9**, 112–145 (1948).

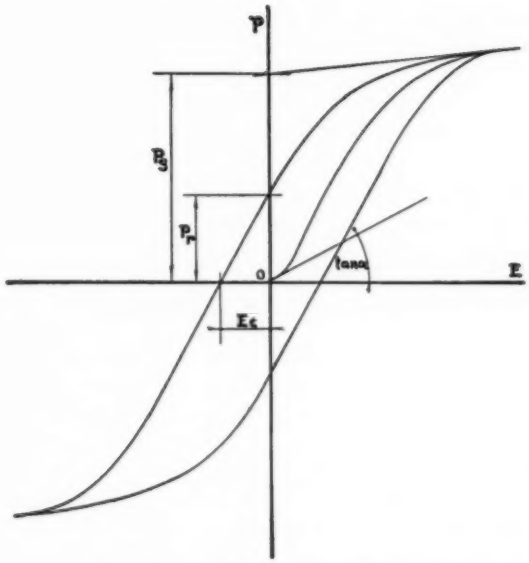


FIG. 1. Hysteresis loop of barium titanate.

Attempts have already been made to get an electrostatically induced permanent memory using electrostatic forces for the storage of information. This was done by placing on an insulator charge patterns corresponding to the information to be stored. By the attracting forces of this electrostatic charge pattern, very small flowing particles could be deposited, which would show a pattern similar to the original change pattern. The whole process can be conceived of as the development of the electrostatic charge patterns by the mutual attraction existing between the charge pattern and the small moving particles.[6,7]

This method suggested, as a new possibility, the developing and fixing of electrostatic charge patterns by the use of movable dipoles instead of movable particles. It appeared that it might be possible to use a ferroelectric material in which the polarization, like the magnetization of a ferromagnetic material, is a function of the applied field, and in which a strong remanent polarization can be obtained. However, the analogy is not perfect, because the ferroelectric substances such as $BaTiO_3$ do not have external fields like magnets or electrets, and the signal can therefore be obtained only by utilizing the piezoelectric effect by applying mechanical stresses such as ultrasonic waves.

PERTINENT PROPERTIES OF THE BARIUM TITANATES[8]

Ferroelectricity, the spontaneous alignment of permanent electric dipoles by an applied external field, can be considered as stored potential energy, which increases the internal energy of the crystals. This energy may be expressed simply, according to Debye,[9] as

$$Q_p = -\tfrac{1}{2}\alpha N E^2 = -\tfrac{1}{2}\beta P^2.$$

This expression, gives the change in energy per unit volume due to spontaneous polarization, where α=polarizability, E=intensity of electric field, β=Lorenz factor. This energy, also called activation energy, would be of the order of magnitude of 50 watt-sec/gram if total polarization could be achieved in a single crystal, while in a polycrystalline ceramic structure only 10 percent of this value may be obtained. The polarization P can be calculated by the product of the number of dipoles, N, per unit volume and their average dipole moment, $\bar{\mu}$. (One cm³ contains 1.56×10^{22} complete cells with an average dipole moment, $\bar{\mu}$, 4.34×10^{-18} esu, or simply 4.34 Debye.)

Figure 1 is a plot of polarization P *versus* electric field E. Several significant points noted on the curve are:

Total polarization, P_t, is the theoretical maximum polarization obtainable if all the dipoles were to be aligned in one direction.

Remanent polarization, P_r, is the polarization remaining when the electric field E is reduced to zero. Just as the magnetic hysteresis loop can today be varied by pretreatment and alloying, the electric remanence and coercive force can also be controlled. (For example, by introducing some lead titanate, a remanent polarization that is really permanent may be produced.)

Coercive field, E_c, is the field needed for the obliteration of the remanent polarization P_r. The actual value of the coercive field is of the order of 25,000 volt/cm.

Spontaneous polarization, P_s, shows the alignment of permanent electric dipoles caused by an applied external field, and it is determined by the highest applicable field strength. This is the most characteristic property of a ferroelectric material. The actually obtained value of P_s is about 15.5 microcoulombs/cm² or 46,500 esu.

Prepolarization may be any degree of orientation of the normally randomly distributed permanent electric dipoles.

Piezoelectric property.—If $BaTiO_3$ material is polarized in any degree, piezoelectric properties are induced in it.

Figure 2 is a plot of relative piezoelectric sensitivity *versus* time and field parameter, and shows that the saturation polarization is reached sooner with higher potential gradients, but that even small voltages polarize the material ultimately.[10]

Figure 3 is a plot of spontaneous polarization, P_s, *versus* temperature.[11] The quantity P_s is the value of polarization in the (001) direction, which is measured

[6] Paul Selenyi, Z. tech. Physik No. 12 (1936).
[7] R. M. Schaffert, Science News Letter (October 23, 1948).
[8] A. von Hippel, Revs. Modern Phys. **22**, 221 (1950).
[9] P. Debye, *Polar molecules* (The Chemical Catalog Company, Inc., New York).
[10] Howatt, Grownover, and Dramets, Electronics **21**, No. 12 (December, 1948).
[11] W. J. Mertz, Technical Report XIV, Laboratory for Insulation Research, M.I.T., 1949; Phys. Rev. **76**, 1221 (1949).

directly by the charge accumulated on electrodes applied to (001) faces of a single crystal.

Figure 3 shows that when a $BaTiO_3$ sample cools and passes into the ferroelectric range below the Curie temperature, a great jump in spontaneous polarization occurs. In other words a ΔT of a few degrees just below the Curie temperature causes the randomizing action of thermal motion to vanish very rapidly. This fact is most significant, since it is possible to cause a fast randomizing action in a previously oriented $BaTiO_3$ material. By a similar process, previously randomized permanent electric dipoles may be easily polarized. It was observed that the transition is of a second-order type with no latent heat; but with a very small jump in specific heat, the influence of which can be neglected in this discussion since the change of the specific heat, Δc, near the Curie point was observed to be around 3.10^3 cal per gram per degree, while $\int_{T_0}^{T_c}\Delta c dT$, the total energy represented by the specific heat anomaly, was found[12] to be approximately 0.25 watt-sec/gram which is only approximately one two-hundredth part of the obtainable polarization energy.

Writing and erasing energy considerations.—It may be seen from Figure 3 that in the vicinity of the Curie point the orientation or randomizing action can be accomplished by a small change of energy. It should be noted that if ΔT is around 3°C, the drop of saturation polarization is not greater than 50 percent. In other words, at 3°C below the Curie temperature 50 percent of the randomizing action of thermal motion has already vanished. For the sake of simplicity, this abrupt jump of polarization may be considered to be linear when ΔT does not exceed 10°C. [For practical purposes, all energies will hereafter be expressed in watt-seconds and will be referred to a unit volume of one cubic centimeter or one gram.]

The process of storing information in a ferroelectric material may be accomplished at a working temperature of a few degrees below the Curie point by using electric signals, representing the information, to supply the small additional energy needed to cause a polarization pattern by randomizing a prepolarized state in a ferroelectric dielectric.

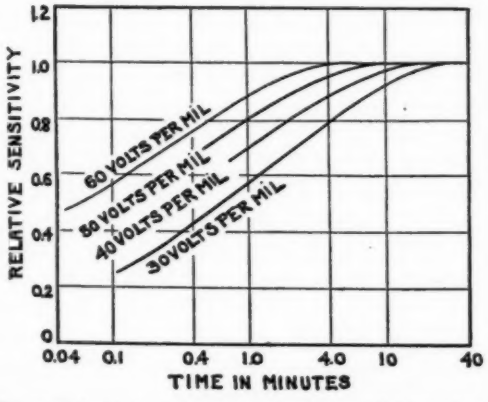


FIG. 2. Polarization time as a function of field strength.

[12] Wal, Blattner and Merr, Harwood, Popper, and Rushman.

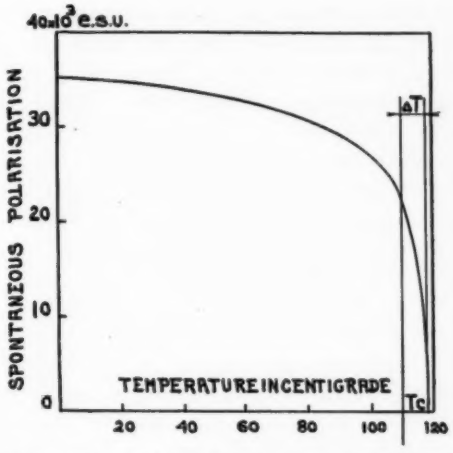


FIG. 3. Spontaneous polarization as a function of temperature.

The following brief practical calculation is made to show the order of magnitude of writing and erasing energy. Suppose that we are working at ΔT degrees below the Curie point. Then the necessary additional randomizing heat energy for reaching the Curie point may be expressed by the formula

$$Q_H = c\rho V\Delta T; \tag{1}$$

substituting the actual values, we obtain for $Q_H = 1.4$ cal/cm³/°C.

This means that the residual polarization, P, at $(T_c - \Delta T)$ (which is about one-half of the maximum polarization, P_s, of 46,500 esu) can be obliterated in the given case by the addition of the small energy of $Q_H = 6\Delta T$ watt-sec/cm³.

From the previous reasoning, it appears that in the case of writing and erasing the principal problem is to introduce a small additional energy into a prepolarized ferroelectric dielectric. This energy may be introduced by different methods. For instance, writing may be accomplished by pulsed or modulated high frequency, in which case the pulses or modulation represents the information, while erasing may be done by simply heating or high frequency heating.

We have two ways for electrically affecting a condenser: by using either dc or ac. The ac may consist of either pulses, modulated high frequency, or pulsed high frequency. In principle, it does not matter whether cathode rays or simply conductive connections are used.

The electrostatic energy of a dielectric caused by a field can be expressed by the equation

$$Q_s = \epsilon_c(E_p^2/8\pi)V. \tag{2}$$

Comparing Eqs. (1) and (2), we see that if

$$Q_s/Q_H \geqq 1, \tag{3}$$

proper conditions for a complete change in the state of polarization may be obtained.

For a field strength of 20,000 volts per centimeter (chosen to be slightly below the ultimate dielectric strength of barium titanate), taking a ΔT of 1°C, and taking $\epsilon = 6000$ as an average value around the Curie point, the energy-ratio Eq. (3) is not satisfied, since $Q_s/Q_H = 0.018$ for these conditions.

If the expressions for Q_s and Q_H given in Eqs. (1) and (2) are substituted in Eq. (3), it is evident that the energy ratio is independent of the volume, and, therefore, this ratio cannot be improved by simply using thinner layers of dielectric.

As shown above, dc operation does not seem very promising, since the electric field required for obtaining a proper energy ratio would be in the order of several hundred kilovolts which far exceeds the breakdown voltage of a $BaTiO_3$ ceramic.

However, this energy ratio could be greatly improved if a very thin layer were used and if, in addition to the energy of the electrical charge, use were to be made of the conductivity effect induced by electron bombardment.[13] In any case, breakdown of the dielectric will limit the power which may be applied.

The following calculation shows that high frequency may be used to supply the required additional energy in a very short time to a condenser, and in this way the aforementioned energy ratio can be arbitrarily increased. The induced heat energy in a condenser depends on the applied electrical energy multiplied by the loss factor of the dielectric, and therefore a higher loss factor results in a higher heat energy introduction rate. We know already that $BaTiO_3$ has a high relaxation frequency and that $\tan\delta$ increases rapidly in the range of 10~100 mc. This increase is as high as approximately ten times the value observed in the kc range. Taking into account that at higher frequencies the impedance of the same condenser decreases while the applied field may be kept constant, it appears that very high writing speed may be expected.

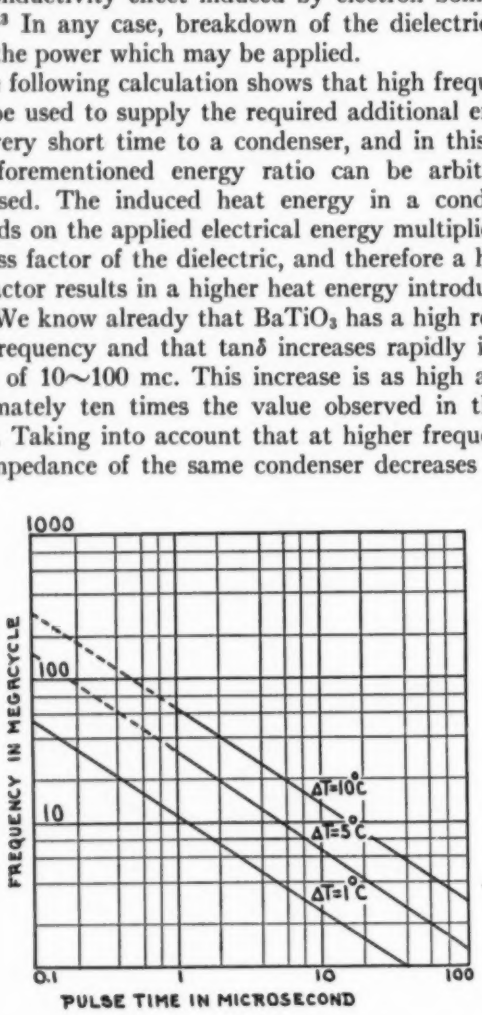


FIG. 4. Pulse duration required for inscribing process as a function of temperature increment and frequency of applied pulses.

An expression for the heat energy developed by a high frequency pulse may be derived from Eq. (2).

$$W_p = (5/9)\nu\epsilon \tan\delta E_p^2 10^{-12} t \text{ watt-sec/cm}^3. \quad (4)$$

Now, E_p will represent the rms value of the high frequency voltage. Again forming the energy ratio of Eq. (3), taking the limiting case where $W_p = Q_H = 6\Delta T$ watt-sec/cm³, and solving for ν, the minimum required frequency of the pulse as a function of t, the minimum required duration of the pulse (expressed in microsec), the following relation is obtained:

$$\nu = (9Q_H 10^{12})/(5\epsilon E_p^2 \tan\delta t) \text{ cycles/sec.} \quad (5)$$

Where Q_H, ϵ, and E_p are as before, and $\tan\delta$ is taken from a table prepared by Warren P. Mason.[14] Although Mason's table refers to single crystals rather than ceramics, there is an approximate agreement with values given elsewhere[5] for $\tan\delta$ at 10 kc. Then, too, this work was done at a field strength of 1.181 volt/cm rather than 25,000 volt/cm; however, since no other values of $\tan\delta$ around the Curie point are available, these have been used for a first approximation. Table II is reproduced from Mason's work.

TABLE II. $\tan\delta$ of $BaTiO_3$ as a function of applied frequency.

ν	10×10^3	50×10^3	100×10^3	500×10^3	1×10^6	5×10^6	10×10^6	50×10^6
$\tan\delta$	0.05	0.05	0.05	0.075	0.1	0.22	0.4	0.5

The function $\nu = f(t)$ was evaluated by successive approximations for $\Delta T = 1°$, 5°, and 10°. A graph of this function is given in Fig. 4. This diagram gives an approximate picture of ac operation under the given conditions. The appropriate frequency for any given pulse time may either be computed from Eq. (5) or estimated from the graph in Fig. 4. The thickness of the ferroelectric affects only the power requirements, and therefore may be chosen on the basis of practical considerations.

It is now apparent that a memory based on these principles will require the development of considerable high frequency technique. From these considerations it appears that records up to the Mc band may be quite possible, or that a lower temperature ($T_c - \Delta T$) may be chosen. In the case of a material with a low T_c, inscribing by a modulated high frequency and erasing also may be done at room temperature.

Generally speaking, any type of radiation which has a randomizing or heating effect may cause a permanent pattern or path in a prepolarized ferroelectric material. For instance, since nuclear kinetic energy is in fact also heat energy, it is likely that this type of electrostatic memory may become an important new research tool for detecting the path of high energy particles in solids.

[13] L. Pensoc, Phys. Rev. **75**, 972–978 (1949).

[14] Warren P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (D. Van Nostrand Company, Inc., New York, 1950).

To summarize, the operating cycle of a feroelectric memory might be as follows:

1. Uniform prepolarization by the application of a dc field for an extended time, possibly hastened by cooling through the Curie point.
2. Local depolarization, in a pattern corresponding to the information to be recorded, by means of modulated high frequency, with preheating, if necessary.
3. Reading, as often as desired, with ultrasonics.
4. Erasing by heating above Curie point by high frequency energy or otherwise.

This cycle may be repeated indefinitely, or may be stopped at step 3.

EXPERIMENTAL SETUP

An experiment was performed to demonstrate that high frequency can, indeed, be used to reduce or eliminate polarization in ferroelectric ceramics.

Figure 5 shows the experimental setup, which consisted essentially of a series of condensers with a dielectric of $BaTiO_3$, 0.14 mm thick. Each of the condensers had a surface area of 1 mm². One of the plates was common and grounded, while the other plates were separate and connected to a double switch system, through which the high frequency pulses could be applied or a connection could be produced to the input of an amplifier and oscilloscope. The condensers are placed in a thermostatically controlled chamber. For measuring the residual polarization, the condensers were acted upon by ultrasonic waves produced by an ultrasonic generator, H.F.G_2., and the piezoelectric transducer. As indicated in the diagram, both the base plate and the working plate were made $\lambda/2$ thick for the ultrasonic waves in order to achieve best mechanical coupling. The individual condensers are formed on a common $BaTiO_3$ sheet which is attached by a heat resistant cement.

The experiment was carried out as follows: after heating to 114°C, the individual condensers were acted upon, one after another, at intervals of 15 seconds. The cooling commenced immediately after the last condenser was acted upon. The residual polarization was measured in relative units. The applied ultrasonic pressure was approximately 800 dynes/cm².

EXPERIMENT WITH 205 kc

For my first experiment, I chose a pulsed high frequency of 205,000 cycles at 225 volts. The pulse time duration ranged from 3×10^{-4} sec up to 5×10^{-2} sec. The data obtained are summarized in Fig. 6.

Figure 6 is a three-dimensional plot. The abscissa gives the pulse time t of the applied high frequency pulses, the ordinate gives the residual polarization of P_r in an arbitrary scale, which has been measured by ultrasonic agitation. Axis z is a parameter showing the time which has elapsed between the application of the high frequency pulse and the cooling of the samples from the working temperature.

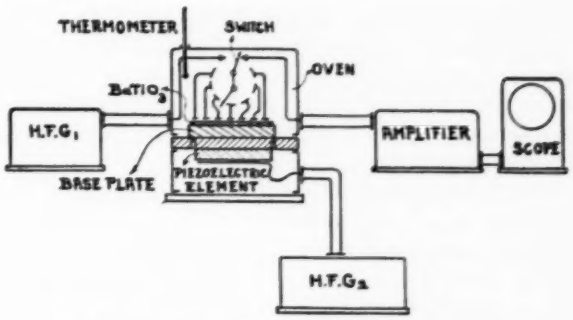


FIG. 5. Setup of experimental ferroelectric memory.

Each point on the diagram represents an average of four readings, and all measurements were repeated several times. The pulse time t was also taken as an average of *ca* 10 measurements. All the curves were taken for a working temperature 3°C below the Curie point, i.e., in this particular case, at 114°C. In the front $t''=0$ plane of the plot is a curve that shows a decrease of residual polarization as a function of the applied pulse time t when the sample is cooled immediately after the high frequency treatment. This curve may represent a type of "frozen-in" change and will be referred to as curve $(t''=0)$.

The curves marked with 15, 30, and 45 seconds, were taken by cooling the samples at those intervals after the high frequency treatment. These represent states in which the residual polarization, represented by the curve $(t''=0)$, has had an opportunity to regroup the randomized but easily moving dipoles. Referring particularly to the curve $(t''=0)$, we may differentiate two sections; the first of which may be represented by ABCD and which reaches to *ca* 3×10^{-4} sec. This section shows the greatest change in polarization. A drop of the prepolarized state as great as 70 percent may be noticed, and this indicates that at that point the applied additional energy of the pulse caused the material to reach the Curie point.

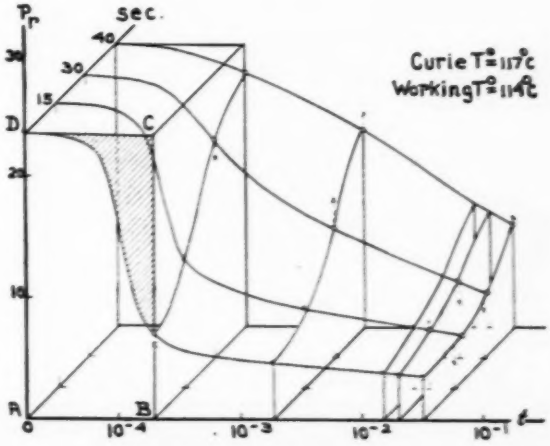


FIG. 6. Residual polarization as a function of pulse duration and cooling time.

The second section which reaches up to 5×10^{-2} sec represents that range in which the applied additional electrical energy was sufficient to exceed the Curie point. This range may be used for obtaining yes and no patterns, which would resemble the punched-card type of memory device.

The discrepancies between experiment and calculation may be accounted for by the extreme value of ϵ and $\tan\delta$ around the Curie point.

This short investigation, which is summarized in Fig. 6, indicates that the use of modulated or pulsed high frequency in the kc range clearly demonstrates the possibility of building memories as fast as may be needed for recording in the sound spectrum.

EXPERIMENTS AT 150 Mc

A surplus radar transmitter was used as a source of 7-microsec pulses at 150 Mc. The transmitter had provisions for varying the peak power of the pulses up to a maximum of 1 kw. A quarter-wave transformer and a tuning stub were used to match the transmitter to the impedance of the piezoelectric condensers and to tune out their capacity.

The experiment was performed as follows: Four piezoelectric condensers were given approximately equal initial polarizations. These condensers were then heated to 114°C ($\Delta T=3°$), acted upon by a single 7-microsec, 150 Mc, 0.5-kw pulse, and immediately cooled to room temperature. The remaining polarization was again measured and found to average around one-third of the initial polarization.

The described experimental setup was not suited for quantitative measurements in the ultra high frequency range, and due to this fact the data of Fig. 4 and the results obtained with 150-Mc pulses are not directly comparable.

This work was done as an investigation preliminary to the undertaking of a research project on the development of different types of memory devices using the described principles.

ACKNOWLEDGMENT

The work here reported was started with the financial support of the U. S. Navy Bureau of Ordnance (contract NOrd 9951) and is being continued under contract with the USAF Office of Air Research (AF33(038)15977). The author wishes to thank Dr. B. D. Van Evera for his kind cooperation and Dr. Leslie S. G. Kovasznay (The Johns Hopkins University) for his helpful criticism.

LIST OF SYMBOLS

α = polarizability.
β = Lorenz factor or dielectric impermeability for the first approximation.
C = capacity of a condenser.
c = specific heat of $BaTiO_3$ (taken as 0.26 gram cal/cm³).
Δc = specific heat anomaly.
$\tan\delta$ = loss factor of dielectric.
ϵ = dielectric constant of material.
ϵ_c = average dielectric constant around T_c, taken as 6000.
E = intensity of electric field.
E_c = coercive field.
E_p = electric potential, volt/cm, field strength.
N = number of dipoles per unit volume.
P = polarization.
P_r = remanent polarization.
P_s = spontaneous polarization.
P_p = permanent polarization.
Q_p = energy of polarization.
Q_H = heat energy.
t = pulse duration time.
t'' = heating time parameter.
T = temperature in ° centigrade.
T_c = Curie temperature.
ΔT = deviation of temperature.
V = volume.
Q_s = electrical energy of a charged condenser.
W_p = electric energy of high frequency pulse.
$\bar{\mu}$ = average dipole moment.
ρ = density of $BaTiO_3$, 5.5 gram/cm.
ν = frequency of electrical writing energy.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

Elastic Energy of an Ellipsoidal Inclusion in an Infinite Solid*

KENNETH ROBINSON
Columbia University, New York, New York
(Received March 7, 1951)

An exact solution in closed form is presented for the displacements, stresses and total elastic energy for the problem of an ellipsoidal isotropic inclusion bonded to an isotropic body of infinite extent when the inclusion and the infinite solid have uniformly distributed changes in their stress-free specific volumes. The physically important details of the dependence of the total strain energy of the two materials upon the shape and volume of the inclusion, the elastic moduli of the two isotropic materials and the changes of specific volume, are included. The solution is indicated for the stresses and displacements for the case where the material at infinity is, in addition to the above condition, in a homogeneous state of stress whose principal directions are parallel to the axes of the cavity.

NOMENCLATURE

$\mathbf{r}$ = position vector.
$\mathbf{u}, \mathbf{u}'$ = displacement vectors for exterior and interior, respectively.
Ψ, Ψ' = stress dyadics for exterior and interior, respectively.
G, G' = shear moduli for exterior and interior, respectively.
ν, ν' = Poisson's ratio for exterior and interior, respectively.
$\epsilon T, \epsilon' T$ = dimensionless products of the coefficient of expansion and temperature changes for exterior, and interior, respectively.
$\sigma_1, \sigma_2, \sigma_3; e_1, e_2, e_3$ = principal stresses and strains at infinity.
$\sigma_1', \sigma_2', \sigma_3'; e_1', e_2', e_3'$ = principal stresses and strains in inclusion.
a, b, c = semiaxes of ellipsoidal inclusion $a > b > c$.
$\rho_1 = b/a, \rho_2 = c/b$ = shape ratios of ellipsoidal inclusion.

1. INTRODUCTION

SOLUTIONS of the equations of elasticity pertaining to the effects of change of specific volume of inclusions of various shapes and elastic properties have been considered by a number of authors. Goodier[1] has reduced to a standard problem in potential theory the general case of an inclusion of arbitrary shape embedded in an infinite matrix subject to the restriction that the inclusion and the matrix have the same elastic properties. Applications of Goodier's theory to the spheroidal and semi-infinite cylindrical inclusions were made by Mykelstad.[2] Mindlin and Cheng[3] extended Goodier's method to the semi-infinite matrix and solved the problem of the spherical inclusion. Except for the simple problems of the sphere and circular cylinder in the infinite matrix, there appear to be only two exact solutions for the case where the inclusion and the matrix have different elastic properties. These are by Mindlin and Cooper[4] for the elliptic cylinder, and Edwards[5] for the spheroidal inclusion. Nabarro[6] has made approximate studies and has drawn qualitative conclusions in connection with the influence of the shape of the inclusion on the elastic energy. The main purposes of this paper are to present the exact solution for the ellipsoidal inclusion and to draw quantitative conclusions regarding the influence of its shape on the elastic energy.

We consider an ellipsoidal inclusion bonded (i.e., continuous displacement and tractions across the interface) to a body of infinite extent. The inclusion, henceforth called the interior, and the surrounding body, henceforth called the exterior, are of different homogeneous isotropic elastic materials whose elastic moduli do not change with temperature. It is assumed that the bodies are under no initial stress when at a uniform common temperature, which is assumed zero for convenience. The temperature of the two bodies is changed, within the limits of linear elasticity, to a uniform temperature T so that the stress free strains of the interior and exterior are $\epsilon' T$ and ϵT respectively. (ϵ', ϵ are the coefficients of linear expansion.) If the interior and the exterior are subjected to uniform changes in their stress-free specific volumes, such that the above given stress free strains are present, the problem is identical. An exact solution in closed form for the displacements, stresses and total elastic strain energy of the two bodies, together with a discussion of the physically important

* Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Joint Committee on Graduate Instruction, Columbia University, New York, New York.

[1] J. N. Goodier, Phil. Mag., Ser. 7, **23**, 1017–1032 (1937).
[2] N. O. Mykelstad, J. Appl. Mech. **9**, A136–143 (1942).
[3] R. D. Mindlin and D. H. Cheng, J. Appl. Phys. **21**, 931–933 (1950).
[4] R. D. Mindlin and H. J. Cooper, J. Appl. Mech. **17**, 265–268 (1950).
[5] R. H. Edwards "Stress concentration around spheroidal inclusions and cavities," Paper No. 50-APM-20, presented at the Annual Conference of the Applied Mechanics Division of the Am. Soc. Mech. Engrs., Purdue University, Lafayette, Indiana, (June 1950).
[6] F. R. N. Nabarro, Proc. Phys. Soc. London **52**, 90–93 (1950). F. R. N. Nabarro, Proc. Roy. Soc. (London), Series A, **175**, 519–538 (1940).

details of the strain energy are presented in this paper. The solution of the more general problem of the infinite material subjected to uniform tensions at infinity in addition to the above conditions, is also indicated.

It was noted by Mindlin and Cooper[4] that the stress in the interior of the elliptic cylinder was homogeneous when the cylinder and matrix were subjected to a homogeneous change in temperature. The assumption, later verified, of homogeneity of stress in the interior of the ellipsoid provided a means for generalizing the results of Sadowsky and Sternberg.[7] Somewhile after the problem was solved, the publication of Edwards' paper[5] showed this same homogeneity of stress to exist in the spheroidal inclusion.

The cartesian coordinate axes are chosen coincident with the principal axes of the ellipsoid whose surface is then given by

$$x^2/a^2+y^2/b^2+z^2/c^2=1, \tag{1.01}$$

where a, b, c, satisfy the relations $a>b>c>0$.

In the absence of body forces the thermoelastic equations of equilibrium in terms of displacement for both regions are

$$(1-2\nu)\nabla^2\mathbf{u}+\nabla\nabla\cdot\mathbf{u}=2(1+\nu)\nabla(\epsilon T) \tag{1.02}$$

$$(1-2\nu')\nabla^2\mathbf{u}'+\nabla\nabla\cdot\mathbf{u}'=2(1+\nu')\nabla(\epsilon' T), \tag{1.03}$$

with the stress and strain dyadics related by

$$\Psi=G[\nabla\mathbf{u}+\mathbf{u}\nabla]+[G/1-2\nu][2\nu\nabla\cdot\mathbf{u}-(1+\nu)\epsilon T]\mathbf{I}, \tag{1.04}$$

$$\Psi'=G'[\nabla\mathbf{u}'+\mathbf{u}'\nabla]+[G'/1-2\nu'][2\nu'\nabla\cdot\mathbf{u}'-(1+\nu')\epsilon' T]\mathbf{I}, \tag{1.05}$$

where $\mathbf{I}$ is the idemfactor and

$$\left.\begin{aligned}\Psi&=\sigma_x\mathbf{ii}+\sigma_y\mathbf{jj}+\sigma_z\mathbf{kk}+\tau_{yz}(\mathbf{jk}+\mathbf{kj})\\&\qquad+\tau_{zx}(\mathbf{ki}+\mathbf{ik})+\tau_{xy}(\mathbf{ij}+\mathbf{ji})\\\Psi'&=\sigma_x'\mathbf{ii}+\sigma_y'\mathbf{jj}+\sigma_z'\mathbf{kk}+\tau_{yz}'(\mathbf{jk}+\mathbf{kj})\\&\qquad+\tau_{zx}'(\mathbf{ki}+\mathbf{ik})+\tau_{xy}'(\mathbf{ij}+\mathbf{ji}).\end{aligned}\right\} \tag{1.06}$$

The traction acting at a point on a surface with normal $\mathbf{n}$ is given by

$$\mathbf{T}=\Psi\cdot\mathbf{n} \tag{1.07}$$

so that for the problem under consideration the boundary conditions are

$$\Psi'\cdot\mathbf{n}'=-\Psi\cdot\mathbf{u} \qquad \text{on ellipsoidal surface,} \tag{1.08}$$

$$\mathbf{u}'=\mathbf{u} \qquad \text{on ellipsoidal surface,} \tag{1.09}$$

$$\Psi=0 \qquad \text{when } |\mathbf{r}|=\infty, \tag{1.10}$$

where $\mathbf{n}'$, $\mathbf{n}$ are the exterior and interior normals of the ellipsoid.

[7] M. A. Sadowsky and E. Sternberg, J Appl. Mech. 16, 149–157 (1949).

2. TOTAL STRAIN ENERGY OF INCLUSION AND INFINITE SOLID[8]

Consider the inclusion and its matrix in their final states (i.e., at a homogeneous temperature T) with a distribution of traction acting on the surface "S" of the inclusion and displacements $(\mathbf{u}-\epsilon T\mathbf{r})$, $(\mathbf{u}'-\epsilon' T\mathbf{r})$ of the exterior and interior, measured from the unstressed final temperature states. The strain energy of the exterior and interior are as follows:

Strain energy of exterior

$$=\frac{1}{2}\int\int_S(\mathbf{u}-\epsilon T\mathbf{r})\cdot\Psi\cdot\mathbf{n}dS \tag{2.01}$$

Strain energy of interior

$$=\frac{1}{2}\int\int_S(\mathbf{u}'-\epsilon' T\mathbf{r})\cdot\Psi'\cdot\mathbf{n}'dS, \tag{2.02}$$

where the region of integration is the surface of the ellipsoid. The total strain energy W, obtained by summing Eqs. (2.01) and (2.02) and then utilizing Eqs. (1.08) and (1.09) as well as the constancy of $(\epsilon T-\epsilon' T)$, is given by

$$W=\tfrac{1}{2}(\epsilon T-\epsilon' T)\int\int_S\mathbf{r}\cdot\Psi'\cdot\mathbf{n}'dS. \tag{2.03}$$

Transformation of the surface integral by Green's theorem and the identity[9]

$$\nabla\cdot(\Psi\cdot\mathbf{V})=(\nabla\cdot\Psi)\cdot\mathbf{V}+\Psi:\Delta\mathbf{V} \tag{2.04}$$

results in

$$W=\tfrac{1}{2}(\epsilon T-\epsilon' T)\int\int\int_V(\nabla\cdot\Psi')\cdot\mathbf{r}+\Psi':\mathbf{I}dV \tag{2.05}$$

where the integration extends over the interior region. In the absence of body force, the equilibrium equation in terms of stress is

$$\nabla\cdot\Psi'=0 \tag{2.06}$$

and the remainder of the integrand is

$$\Psi':\mathbf{I}=\sigma_1'+\sigma_2'+\sigma_3'=\theta' \tag{2.07}$$

so that the total elastic energy is given by

$$W=\tfrac{1}{2}(\epsilon T-\epsilon' T)\int\int\int_V\theta' dV. \tag{2.08}$$

When the sum of the principal stresses θ' (first invariant of the stress dyadic) is constant in the interior, the total strain energy is given by

$$W=\tfrac{1}{2}(\epsilon T-\epsilon' T)\theta' V \tag{2.09}$$

where V is the volume of the inclusion.

[8] The development of this section is essentially that given by W. Thornton Read, Jr. in a private communication.

[9] For Green's theorem and the identity (2.04), see C. E. Weatherburn, *Advanced Vector Analysis* (G. Bell and Sons, Ltd., London, 1947).

3. GENERAL SOLUTIONS IN ORTHOGONAL CURVILINEAR COORDINATES

Sadowsky and Sternberg[7] have considered a homogeneous, isotropic, elastic body of infinite extent which possesses a cavity in the shape of an ellipsoid, assuming that the body at infinity is in a uniform state of stress whose principal axes of stress are parallel to the axes of the cavity. They made use of the following general solution to Eq. (1.02), due to J. Boussinesq,[10] for the case of a uniform temperature in the exterior:

$$2G\mathbf{u}=\nabla(\mathbf{r}\cdot\mathbf{B}+F)-4(1-\nu)\mathbf{B} \tag{3.01}$$

where

$$\mathbf{B}=X\mathbf{i}+Y\mathbf{j}+Z\mathbf{k}$$

and

$$\nabla^2\mathbf{B}=\nabla^2F=0. \tag{3.02}$$

In order to describe the ellipsoidal boundary of the cavity with a single coordinate, they utilize an orthogonal curvilinear coordinate system with the transformation,

$$x=kms_1s_2s_3,\quad y=-kmc_1c_2c_3/k',$$

$$z=imd_1d_2d_3/kk' \tag{3.03}$$

where $i=(-1)^{\frac{1}{2}}$ and the elliptic functions, their moduli k, k', and the parameter m are given by

$$\left.\begin{array}{lll} k=(a^2-b^2/a^2-c^2)^{\frac{1}{2}} & k'=(1-k^2)^{\frac{1}{2}} & m=(a^2-b^2)^{\frac{1}{2}} \\ s_1=1/ksn\alpha_1 & s_2=1/dn(\alpha_2,k') & s_3=sn\alpha_3 \\ c_1=dn\alpha_1/iksn\alpha_1 & c_2=-ik'sn(\alpha_2,k')/dn(\alpha_2,k') & c_3=cn\alpha_3 \\ d_1=cn\alpha_1/isn\alpha_1 & d_2=k'cn(\alpha_2,k')/dn(\alpha_2,k') & d_3=dn\alpha_3. \end{array}\right\} \tag{3.04}$$

Fixed values of the coordinates α_1, α_2, α_3, represent members of the families of confocal ellipsoids, confocal hyperboloids of one sheet and confocal hyperboloids of two sheets respectively. The surface of the ellipsoidal inclusion indicated by Eq. (1.01) corresponds to a value of $\alpha_1=\alpha_1^0$ evaluated from

$$s_1^0=1/k\ \mathrm{sn}\alpha^0{}_1=a/m. \tag{3.05}$$

Unit normals to the surfaces α_1, α_2, α_3=constant at the point $(\alpha_1, \alpha_2, \alpha_3)$ are given by

$$\mathbf{a}=h_1\nabla\alpha_1,\quad \mathbf{b}=h_2\nabla\alpha_2,\quad \mathbf{c}=h_3\nabla\alpha_3 \tag{3.06}$$

where the local scale coefficients h are obtained from

$$\left.\begin{array}{lll} h_1=i/kmq_2q_3 & h_2=1/kmq_3q_1 & h_3=i/kmq_1q_2 \\ q_1=(s_2^2-s_3^2)^{\frac{1}{2}} & q_2=i(s_1^2-s_3^2)^{\frac{1}{2}} & q_3=(s_1^2-s_2^2)^{\frac{1}{2}} \end{array}\right\}. \tag{3.07}$$

The gradient of a function then becomes

$$\nabla U=\mathbf{a}h_1\partial U/\partial\alpha_1+\mathbf{b}h_2\partial U/\partial\alpha_2+\mathbf{c}h_3\partial U/\partial\alpha_3 \tag{3.08}$$

and the angle between the normals to the coordinate surface α_j, x evaluated from

$$\cos(\alpha_j,x)=h_j\partial x/\partial\alpha_j=(1/h_j)\partial\alpha_j/\partial x. \tag{3.09}$$

The displacements vectors obtained from Eqs. (3.01), (3.09), (3.08) are given by

$$\left.\begin{array}{l} 2G\mathbf{u}=h_1[x\partial X/\partial\alpha_1-(3-4\nu)X\partial x/\partial\alpha_1]\mathbf{a}+h_2[x\partial X/\partial\alpha_2-(3-4\nu)X\partial x/\partial\alpha_2]\mathbf{b} \\ \qquad\qquad +h_3[x\partial X/\partial\alpha_3-(3-4\nu)X\partial x/\partial\alpha_3]\mathbf{c} \\ \cdot\\ \cdot\\ \cdot\\ 2G\mathbf{u}=\mathbf{a}h_1\partial F/\partial\alpha_1+\mathbf{b}h_2\partial F/\partial\alpha_2+\mathbf{c}h_3\partial F/\partial\alpha_3. \end{array}\right\} \tag{3.10}$$[11]

The components of stress in the orthogonal curvilinear coordinates, obtained from Eq. (1.07), are

$$\tau_{11}=\mathbf{a}\cdot\Psi\cdot\mathbf{a}\quad\cdot\tau_{23}=\mathbf{b}\cdot\Psi\cdot\mathbf{c}. \tag{3.11}$$[11]

Expressions for these components of stress in terms of the displacement potentials X, Y, Z, F may be obtained by utilizing Eqs. (1.04) (3.10). They are given in reference 7 as Eqs. [15], and will henceforth be called Eqs. (3.12).

4. ELLIPSOIDAL HARMONICS

Harmonic functions for the generation of stress fields by means of Eqs. (3.10) and (3.12) may be found from the solutions to Laplace's equation (3.02) in the form of Lame products,

$$U_n^p(\alpha_1,\alpha_2,\alpha_3)=\prod_{j=1}^{j=3}[\lambda_n^p(\alpha_j)\text{ or }\Lambda_n^p(\alpha_j)] \tag{4.01}$$

where n, p denote the degree and species, respectively, of the Lame function of the first kind λ_n^p. The associated second solution Λ_n^p of Lame's equation[12] is given by

$$\Lambda_n^p(\alpha_j)=\lambda_n^p(\alpha_j)\int_0^{\alpha_j}\frac{dt}{[\lambda_n^p(t)]^2} \tag{4.02}$$

so that $\Lambda_n^p(0)=0$ and hence the Lame functions of the second kind in α_1 vanish at $|\mathbf{r}|=\infty$.

[10] For a discussion of general solutions and their interdependence see R. D. Mindlin, Bull. Am. Math. Soc. **12**, 373–376 (June, 1936).

[11] Those formulas not given explicitly may be obtained by cyclic permutation.

[12] For Lame's differential equation see reference 7 or E. T. Whitaker and G. N. Watson, *A Course of Modern Analysis* (University Press, Cambridge, England, 1935), fourth edition.

TABLE I.

Solution No.	Degree n	Species p	$\lambda_n{}^p(\alpha_j)$	$\Lambda_n{}^p(\alpha_1)$	$U_n{}^p(\alpha_1, \alpha_2, \alpha_3)$
1	1	1	s_j	$S_1=[\alpha_1-E(\alpha_1)]s_1$	$X=mS_1s_2s_3$
2	1	$1/1+k^2$	c_j	$C_1=[(k')^2\alpha_1-E(\alpha_1)]c_1+(d_1/s_1)$	$Y=-\frac{m}{k'}C_1c_2c_3$
3	1	$k^2/1+k^2$	d_j	$D_1=(c_1/s_1)-d_1E(\alpha_1)$	$Z=\frac{im}{k^2k'}D_1d_2d_3$
4	0	0	1	α_1	$F_1=km^2\alpha_1$
5	2	$2+\frac{2[(k')^2+k^4]^{\frac{1}{2}}}{1+k^2}$	$l_j=s_j^2-\frac{2}{p(1+k^2)}$	L_1	$F_2=km^2L_1l_2l_3$

$$E(\alpha_1)=\int_0^{\alpha_1} dn^2(t)dt,\quad L_1=\{2k^2\alpha_1+p(1+k^2)[E(\alpha_1)-\alpha_1]\}l_1+2c_1d_1/s_1.$$

5. SOLUTION OF A TRACTION PROBLEM FOR THE INFINITE SOLID BOUNDED BY THE SURFACE $\alpha_1=\alpha_1^0$

We consider a body under the state of stress due to the tractions caused by uniform stress fields acting at infinity ($\alpha_1=0$) and tractions on the surface of the ellipsoid $\alpha_1=\alpha_1^0$. The stress dyadic at infinity and the traction acting on the surface $\alpha_1=\alpha_1^0$ are given as

$$\Psi_\infty=[\sigma_1\mathbf{ii}+\sigma_2\mathbf{jj}+\sigma_3\mathbf{kk}] \qquad \text{at } \alpha_1=0 \qquad (5.01)$$

$$\Psi'\cdot\mathbf{n}=[\sigma_1'\mathbf{ii}+\sigma_2'\mathbf{jj}+\sigma_3'\mathbf{kk}]\cdot\mathbf{n} \quad \text{at } \alpha_1=\alpha_1^0 \qquad (5.02)$$

where $\sigma\mathbf{j}$, $\sigma\mathbf{j}'$ are constants ($j=1, 2, 3$).

The components of traction (τ_{11}, τ_{12}, τ_{13}) acting on the surface $\alpha_1=\alpha_1^0$ at the point (α_1^0, α_2, α_3) are obtained by utilizing Eqs. (5.02), (3.04), (3.03), (3.07) and (3.09) and are as follows:

$$\left.\begin{aligned}
\tau_{11}&=-[1/(k')^2q_2^2q_3^2][\sigma_1'(k')^2c_1^2d_1^2s_2^2s_3^2\\
&\qquad+\sigma_2's_1^2d_1^2c_2^2c_3^2-\sigma_3's_1^2c_1^2d_2^2d_3^2]\\
\tau_{12}&=-[s_1c_1d_1s_2c_2d_2/(k')^2q_1q_2q_3^2]\\
&\qquad\times[\sigma_1'(k')^2s_3^2+\sigma_2'c_3^2-\sigma_3'd_3^2]\\
\tau_{13}&=-[s_1c_1d_1s_3c_3d_3/(k')^2q_1q_2^2q_3]\\
&\qquad\times[\sigma_1'(k')^2s_2^2+\sigma_2'c_2^2-\sigma_3'd_2^2]
\end{aligned}\right\}. \qquad (5.03)$$

Sadowsky and Sternberg[7] have solved the problem above for the case of no traction acting on the surface $\alpha_1=\alpha_1^0$. This was achieved by extending the uniform stress field throughout the entire space and superimposing a solution which removes the stress residuals at α_1^0. A consideration of the properties of the stress fields leads to restrictions imposed on the stress functions X, Y, Z, F. These restrictions are that in the region exterior to the cavity, the stress fields must be regular and at infinity they must vanish. Also they must have the symmetry and periodic structure with respect to α_2, α_3 which the uniform field possesses. Accordingly the Lame functions of the second kind in α_1 and of the first kind in α_2, α_3 are the only ones which may be utilized in Eqs. (4.01). A survey of the linearly independent solutions which possess these properties yields the set of stress functions recorded in Table I.

By means of Eqs. (3.03), (3.04), (3.07), (3.12), and Table I, the stress components may be computed. These are recorded as Eqs. [43] through [47], in reference 9. They will be omitted here but referred to as Eqs. (5.04). In the process of solving their problem, Sadowsky and Sternberg have actually solved two problems which, when superimposed, give the solution of the traction problem we are now considering. The dyadic representing the complete state of stress is given by a linear combination of the stress dyadic of Eq. (5.01), and stress dyadics representing the five linearly independent solutions in Eqs. (5.04).

$$\Psi=\Psi_\infty+\sum_{N=1}^{5} A_N\Psi_N. \qquad (5.05)$$

Equating the traction produced by the stress dyadic Ψ to that given by Eq. (5.02) we obtain

$$(\Psi'-\Psi_\infty)\cdot\mathbf{n}=\sum_{N=1}^{5} A_N\Psi_N\cdot\mathbf{n} \quad \text{on} \quad \alpha_1=\alpha_1^0 \qquad (5.06)$$

where $\mathbf{n}$ is the internal normal to the surface $\alpha_1=\alpha_1^0$. The three equations in the components of traction (τ_{11}, τ_{12}, τ_{13}) are obtained with the aid of Eqs. (5.04), (5.03), (5.02), and (5.01). These equations are identities with respect to the variables α_2, α_3 and produce sixteen compatible linear equations in the coefficients of superposition A_N ($N=1, 2, \cdots 5$). They are reducible to the following five linearly independent equations recorded

in matrix form:

$$\begin{pmatrix} P_1 & P_2 & P_3 & 0 & J \\ T_1 & T_2 & T_3 & 0 & -L_1' \\ s_1^2 & -c_1^2 & -d_1^2/k^2 & 0 & \bar{p} \\ s_1^2 T_1 & -c_1^2 T_2 & -d_1^2 T_3/k^2 & 0 & -l_1 L_1' \\ s_1^4 & c_1^4 & d_1^4/k^4 & 1 & l_1 \bar{p} \end{pmatrix} \begin{Vmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{Vmatrix} = \begin{pmatrix} \dfrac{-kc_1d_1}{s_1} & \dfrac{-ks_1d_1}{(k')^2c_1} & \dfrac{k^5s_1c_1}{(k')^2d_1} \\ ks_1c_1d_1 & \dfrac{-ks_1c_1d_1}{(k')^2} & \dfrac{k^2s_1c_1d_1}{(k')^2} \\ 0 & 0 & 0 \\ ks_1^3c_1d_1 & \dfrac{ks_1c_1^3d_1}{(k')^2} & \dfrac{-ks_1c_1d_1^3}{(k')^2} \\ 0 & 0 & 0 \end{pmatrix} \begin{Vmatrix} \sigma_1-\sigma_1' \\ \sigma_2-\sigma_2' \\ \sigma_3-\sigma_3' \end{Vmatrix}. \tag{5.07}$$

$$\left.\begin{aligned} T_1&=(1-2\nu)[1+2S_1c_1d_1] & P_1&=-\frac{1}{s_1^2}\left[\frac{2\nu S_1}{c_1d_1}+T_1+1\right] \\ T_2&=(1-2\nu)\left[1-\frac{2s_1C_1d_1}{(k')^2}\right] & P_2&=-\frac{1}{c_1^2}\left[\frac{2\nu C_1}{s_1d_1}-T_2-1\right] \\ T_3&=(1-2\nu)\left[1+\frac{2k^2s_1c_1D_1}{(k')^2}\right] & P_3&=-\frac{1}{d_1^2}\left[\frac{2\nu D_1}{s_1c_1}-k^2(T_3+1)\right] \\ p&=2+\frac{2[(k')^2+k^4]^{\frac{1}{2}}}{1+k^2} & J&=\frac{L_1}{s_1c_1d_1}[2k^2l_1+(2-p)(1+k^2)] \\ \bar{p}&=\frac{8(1-p)}{3p} & L_1'&=\frac{dL_1}{d\alpha_1}=\frac{2L_1s_1c_1d_1}{l_1}+\frac{\bar{p}}{l_1}. \end{aligned}\right\} \tag{5.08}$$

All the functions in Eqs. (5.07), (5.08) are to be evaluated at $\alpha_1=\alpha_1^0$ for the ellipsoid whose principal semiaxes are (a, b, c). To facilitate these computations it is expedient to introduce shape parameters ρ_1, ρ_2 defined by the relations

$$0<\rho_1=b/a<1, \quad 0<\rho_2=c/b<1. \tag{5.09}$$

Equations (3.04) and (3.05) can then be written as

$$k=\left(\frac{1-\rho_1^2}{1-\rho_1^2\rho_2^2}\right)^{\frac{1}{2}}, \ \mathrm{sn}\alpha_1^0=(1-\rho_1^2\rho_2^2)^{\frac{1}{2}}. \tag{5.10}$$

The equations (5.07), (5.08), (5.05), determine the state of stress resulting from the boundary conditions indicated in Eqs. (5.01), (5.02). It is worthy of note that for the case $\Psi'=0$ (i.e., $\sigma_1'=\sigma_2'=\sigma_3'=0$), the problem and its solution reduce to the one solved by Sadowsky and Sternberg. For the case $\Psi=\Psi'$ (i.e., $\sigma_1=\sigma_1'$, $\sigma_2=\sigma_2'$, $\sigma_3=\sigma_3'$) the coefficients of superposition vanish and the homogeneous state of stress exists throughout the infinite solid bounded by $\alpha_1=\alpha_1^0$.

6. SOLUTION OF A DISPLACEMENT PROBLEM FOR THE INFINITE SOLID BOUNDED BY $\alpha_1=\alpha_1^0$

We consider next the infinite solid subject to displacements on the boundary $\alpha_1=\alpha_1^0$ and at infinity given by

$$\left.\begin{aligned} \mathbf{u}_\infty&=e_1x\mathbf{i}+e_2y\mathbf{j}+e_3z\mathbf{k}=(\bar{e}_1+\epsilon T)x\mathbf{i} \\ &\quad+(\bar{e}_2+\epsilon T)y\mathbf{j}+(\bar{e}_3+\epsilon T)z\mathbf{k} \quad \text{at } \alpha_1=0 \end{aligned}\right\} \tag{6.01}$$

$$\left.\begin{aligned} \mathbf{u}'_{\alpha_1^0}&=e_1'x\mathbf{i}+e_2'y\mathbf{j}+e_3'z\mathbf{k}=(\bar{e}_1'+\epsilon T)x\mathbf{i} \\ &\quad+(\bar{e}_2'+\epsilon T)y\mathbf{j}+(\bar{e}_3'+\epsilon T)z\mathbf{k} \quad \text{at } \alpha_1=\alpha_1^0 \end{aligned}\right\}, \tag{6.02}$$

where $\bar{e}_j$, $\bar{e}_j'$ $(j=1, 2, 3)$ are constant strains due to the stresses alone at infinity and inside $\alpha_1=\alpha_1^0$ respectively, while e_j, e_j' $(j=1, 2, 3)$ include the free thermal expansion.

By analogy with Sec. 5 we consider the displacements to be the superposition of the state at infinity and the displacements given by the five solutions (which vanish at infinity) indicated by Eqs. (3.10) and Table I. Thus we have

$$\mathbf{u}=\mathbf{u}_\infty+\sum_{N=1}^{5} A_N\mathbf{u}_N. \tag{6.03}$$

We must now satisfy Eq. (6.02) which then becomes

$$\mathbf{u}'_{\alpha_1^0}-\mathbf{u}_\infty=\sum_{N=1}^{5} A_N\mathbf{u}_N \quad \text{on} \quad \alpha_1=\alpha_1^0. \tag{6.04}$$

Utilizing Eqs. (3.10), (3.03), (3.07), (3.09), and

$$\left.\begin{aligned} \mathbf{u}'_{\alpha_1^0}-\mathbf{u}_\infty&=u_1\mathbf{a}+u_2\mathbf{b}+u_3\mathbf{c}=(e_1'-e_1)x\mathbf{i} \\ &\quad+(e_2'-e_2)y\mathbf{j}+(e_3'-e_3)z\mathbf{k} \end{aligned}\right. \tag{6.05}$$

TABLE II.

N	u_1^N	u_2^N	u_3^N
1	$\frac{im}{2Gq_2q_3}[1-(2-4\nu)S_1c_1d_1]s_2^2s_3^2$	$\frac{im}{2Gq_3q_1}[-(2-4\nu)s_1S_1]s_2c_2d_2s_3^2$	$\frac{im}{2Gq_1q_2}[-(2-4\nu)s_1S_1]s_3c_3d_3s_2^2$
2	$\frac{im}{2Gq_2q_3}\left[1+(2-4\nu)\frac{s_1C_1d_1}{(k')^2}\right]c_2^2c_3^2$	$\frac{im}{2Gq_3q_1}\left[(2-4\nu)\frac{c_1C_1}{(k')^2}\right]s_2c_2d_2c_3^2$	$\frac{im}{2Gq_1q_2}\left[(2-4\nu)\frac{c_1C_1}{(k')^2}\right]s_3c_3d_3c_2^2$
3	$\frac{im}{2Gq_2q_3}\left[\frac{(k')^2-(2-4\nu)k^2s_1c_1D_1}{k^4(k')^2}\right]d_2^2d_3^2$	$\frac{im}{2Gq_3q_1}\left[\frac{-(2-4\nu)d_1D_1}{k^2(k')^2}\right]s_2c_2d_2d_3^2$	$\frac{im}{2Gq_1q_2}\left[\frac{-(2-4\nu)d_1D_1}{k^2(k')^2}\right]s_3c_3d_3d_2^2$
4	$\frac{im}{2Gq_2q_3}$	0	0
5	$\frac{im}{2Gq_2q_3}[L_1']l_2l_3$	$\frac{im}{2Gq_3q_1}[2L_1]s_2c_2d_2l_3$	$\frac{im}{2Gq_1q_2}[2L_1]s_3c_3d_3l_2$

we obtain for the orthogonal curvilinear components of the displacement vector $(\mathbf{u}'_{\alpha_1^0}-\mathbf{u}_\infty)$

$$\left.\begin{aligned}
u_1&=\frac{im}{k(k')^2q_2q_3}[(e_1'-e_1)k^2(k')^2s_1c_1d_1s_2^2s_3^2-(e_2'-e_2)k^2s_1c_1d_1c_2^2c_3^2+(e_3'-e_3)s_1c_1d_1d_2^2d_3^2]\\
u_2&=\frac{im}{k(k')^2q_3q_1}[(e_1'-e_1)k^2(k')^2s_2c_2d_2s_3^2s_1^2-(e_2'-e_2)k^2s_2c_2d_2c_3^2c_1^2+(e_3'-e_3)s_2c_2d_2d_3^2d_1^2]\\
u_3&=\frac{im}{k(k')^2q_1q_2}[(e_1'-e_1)k^2(k')^2s_3c_3d_3s_1^2c_1^2-(e_2'-e_2)k^2s_3c_3d_3c_1^2c_2^2+(e_3'-e_3)s_3c_3d_3d_1^2d_2^2]
\end{aligned}\right\}. \tag{6.06}$$

The components of the displacement vectors,

$$\mathbf{u}_N=u_1^N\mathbf{a}+u_2^N\mathbf{b}+u_3^N\mathbf{c}\quad(N=1,2,\cdots5) \tag{6.07}$$

obtained from Eqs. (3.10), (3.03), (3.07), and Table I are recorded in Table II. Equations (6.04) combined with (6.05), (6.06), (6.07), and Table II, provide a system of eight compatible equations in the coefficients of superposition A_N $(N=1,2,\cdots5)$. These equations may be reduced to the following set of five linearly independent equations.

$$\left.\begin{aligned}
&[1-(2-4\nu)S_1c_1d_1]A_1+k^2A_4-L_1'\left[\frac{\bar{p}}{4}\right]A_5=2Gks_1c_1d_1(e_1'-e_1)\\
&[-(k')^2-(2-4\nu)s_1C_1d_1]A_2+k^2A_4+L_1'\left[\frac{\bar{p}}{8}-\frac{(k')^2}{p(1+k^2)}\right]A_5=2Gks_1c_1d_1(e_2'-e_2)\\
&\left[\frac{(k')^2-(2-4\nu)k^2s_1c_1D_1}{k^2}\right]A_3+k^2A_4+L_1'\left[\frac{\bar{p}}{8}+\frac{(k')^2}{p(1+k^2)}\right]A_5=2Gks_1c_1d_1(e_3'-e_3)\\
&s_1^2A_1-c_1^2A_2-\frac{d_1^2}{k^2}A_3+\bar{p}A_5=0\\
&s_1^4A_1+c_1^4A_2+\frac{d_1^4}{k^4}A_3+A_4+l_1\bar{p}A_5=0
\end{aligned}\right\}. \tag{6.08}$$

All the functions appearing in Eqs. (6.08) are to be evaluated at $\alpha_1=\alpha_1^0$. The complete solution of the boundary value problem represented by Eqs. (6.01), (6.02) is then given by Eqs. (6.08), (6.03), and Tables I and II. We note now that Eqs. (5.07) for the traction problem solved in Sec. 5 are compatible with Eqs. (6.08). Thus the ellipsoid $\alpha_1=\alpha_1^0$ becomes stretched into another ellipsoid when the infinite solid is under the stress state whose boundary conditions are given by Eqs. (5.01), (5.02).

7. SOLUTION OF THERMOELASTIC PROBLEM

Considering the homogeneous state at infinity represented by Eqs. (5.01) to be the same as that repre-

sented by Eqs. (6.01), we have for the principal stress-principal strain relations (Eqs. (1.04)) at infinity:

$$\sigma_j=[2G/1-2\nu][\nu(e_1+e_2+e_3)+(1-2\nu)e_j-(1+\nu)\epsilon T] \quad (j=1, 2, 3). \quad (7.01)$$

Considering the homogeneous state in the interior represented by Eqs. (5.02) to be the same as that represented by Eqs. (6.02), we have for the principal stress-principal strain relations in the interior:

$$\sigma_j'=[2G'/1-2\nu'][\nu'(e_1'+e_2'+e_3') +(1-2\nu')e_j'-(1+\nu')\epsilon'T] \quad (j=1, 2, 3). \quad (7.02)$$

We note here that the inclusion need only have the property that when under homogeneous principal stresses σ_1', σ_2', σ_3' oriented along the axes of the ellipsoid, it be deformed into another ellipsoid. For example, a cubic crystal with orientation of the crystalographic axes along the axes of the ellipsoid has this property and its thermoelastic stress strain relations could be substituted for Eqs. (7.02).

For given values of the parameters ρ_1, ρ_2, G, G', ν, ν', ϵT, $\epsilon' T$ and a stress state at infinity given by Ψ_∞, Eqs. (5.07), (6.08), (7.02) represent thirteen compatible equations which can be reduced to eleven independent equations in the eleven unknowns σ_1', σ_2', σ_3', e_1', e_2', e_3', A_N ($N=1, 2, \cdots 5$). The state of displacement for the interior may then be determined with the aid of Eqs. (6.02), while the state of displacement and stress for the exterior may be determined from Eqs. (6.03), (5.05), Table II, and Eqs. (5.04).

We confine our attention now to the technically important case[6] in which the stress state at infinity vanishes (i.e., $\Psi_\infty=0$) so that the strains at infinity, as given by Eqs. (7.01) are

$$e_1=e_2=e_3=\epsilon T. \quad (7.03)$$

Substitution of the expressions for the strains e_1, e_2, e_3, e_1', e_2', e_3' from Eqs. (7.02), (7.03) into Eqs. (6.08) then permits the reduction to eight linearly independent equations in the eight unknowns σ_1', σ_2', σ_3', A_N ($N=1, 2, \cdots 5$). Solution of these equations for the sum of the principal stresses ($\sigma_1'+\sigma_2'+\sigma_3'$) gives

$$\theta'=\frac{-6G(\epsilon'T-\epsilon T)}{\eta+[G(1-2\nu')/G'(1+\nu')]}, \quad (7.04)$$

where

$$\begin{aligned}
&\eta=[f_1m_1+f_2m_2+f_3m_3]/[f_1n_1+f_2n_2+f_3n_3],\ g=\frac{1}{2(1-\nu)}\left[\frac{G}{G'}-1\right]\\
&f_1=\bar{p}+g[\bar{p}T_1+\bar{p}T_3+L_1'l_1]+g^2\left[L_1'T_3s_1^2\left(1-\frac{2k^2}{p(1+k^2)}\right)-\frac{2L_1'T_1d_1^2}{p(1+k^2)}+\bar{p}T_1T_3\right]\\
&f_2=-\bar{p}-g[\bar{p}T_1+\bar{p}T_2+L_1'l_1]-g^2\left[L_1'T_2s_1^2\left(1-\frac{2}{p(1+k^2)}\right)-\frac{2L_1'T_1c_1^2}{p(1+k^2)}+\bar{p}T_1T_2\right]\\
&f_3=0+g[-(k')^2s_1^2T_1-c_1^2T_2+d_1^2T_3]+g^2[(k')^2s_1^2T_2T_3+c_1^2T_3T_1-d_1^2T_1T_2]\\
&m_1=c_1^2[2-2\nu-T_1-3k^2s_1^2]-(k')^2s_1^2[2-2\nu-T_2]\\
&m_2=d_1^2[2-2\nu-T_1-3s_1^2]+(k')^2s_1^2[2-2\nu-T_3]\\
&m_3=-\bar{p}[2-2\nu-T_1-6k^2s_1^2/p(1+k^2)]\\
&n_1=[c_1^3d_1S_1+s_1^3d_1C_1](2+2\nu)+3d_1^2[c_1^2-s_1^2]+[c_1^2-(k')^2s_1^2](1-2\nu)\\
&n_2=[c_1d_1^3S_1+k^2s_1^3c_1D_1](2+2\nu)+3c_1^2[d_1^2-k^2s_1^2]+[d_1^2+(k')^2s_1^2](1-2\nu)\\
&n_3=\bar{p}[-c_1d_1S_1(2+2\nu)+\{12k^2s_1^2/p(1+k^2)\}-4+2\nu]
\end{aligned} \quad (7.05)$$

8. THE ELASTIC ENERGY

Since the sum of the principal stresses is a constant in the interior, we may utilize Eq. (2.09) to determine the total elastic energy of the interior and exterior. After substituting θ' from Eq. (7.04) into Eq. (2.09) and utilizing the relations $\rho_1=b/a$, $\rho_2=c/b$, $V=4\pi abc/3$, $G=(G/G')G'$ we may indicate the dependence by the equation,

$$W=\frac{3G'(\epsilon'T-\epsilon T)^2V}{[\eta G'/G]+[(1-2\nu')/(1+\nu')]}, \quad (8.01)$$

where $\eta=\eta(\rho_1, \rho_2, G/G', \nu)$ is given by Eqs. (7.05).

Thus the strain energy depends upon the 8 independent parameters ρ_1, ρ_2, V, G', G/G', ν, ν', $(\epsilon'T-\epsilon T)$.

Special Case No. 1.—For an elliptic disk, i.e., ($\rho_2=0$), the function η becomes independent of ρ_1, ν and is given by

$$\eta(\rho_1, 0, G/G', \nu)=\tfrac{1}{2}G/G',$$

hence,

$$W_1=2G'(1+\nu)(\epsilon'T-\epsilon T)^2V/(1-\nu'). \quad (8.02)$$

Special Case No. 2.—For identical shear moduli of

the interior and exterior regions (i.e., $G/G'=1$), the function η becomes independent of the shape ratios ρ_1, ρ_2 and the exterior Poisson's ratio. It reduces to the constant value $\eta=\frac{1}{2}$ so that the function W is identical with that for the elliptic disk (i.e., W_1). Under the further assumption that $\nu=\nu'$ the value of W given by Eq. (8.02) reduces to that obtainable from Goodier's solution[1] for the most general possible shape inclusion where $G=G'$, $\nu=\nu'$ and the temperature in the inclusion is uniform.

Special Cases Nos. 3 and 4.—For an elliptic cylinder (i.e., $\rho_1=0$) the function η is algebraic and is given by

$$\eta_3=\frac{\{G/G'\}\{(6-8\nu)\rho_2+G/G'[(2-2\nu)(1+\rho_2^2)-(2-4\nu)\rho_2]\}}{(3-4\nu)\rho_2+[G/G'][(4-4\nu)(1+\rho_2^2)+(2-4\nu)\rho_2]+(G/G')^2(3\rho_2)}. \tag{8.03}$$

Under the further assumption that $\rho_2=1$ (i.e., circular cylinder), the function η reduces to

$$\eta_4=2G/(G'+3G) \tag{8.04}$$

and this value is readily verified from Lame's solution[13] for a pressure applied to a concentric cylindrical hole in a circular cylinder.

Special Case No. 5—For a sphere (i.e., $\rho_1=\rho_2=1$) the value of η reduces, as in case No. 2, to $\eta=\frac{1}{2}$ and the

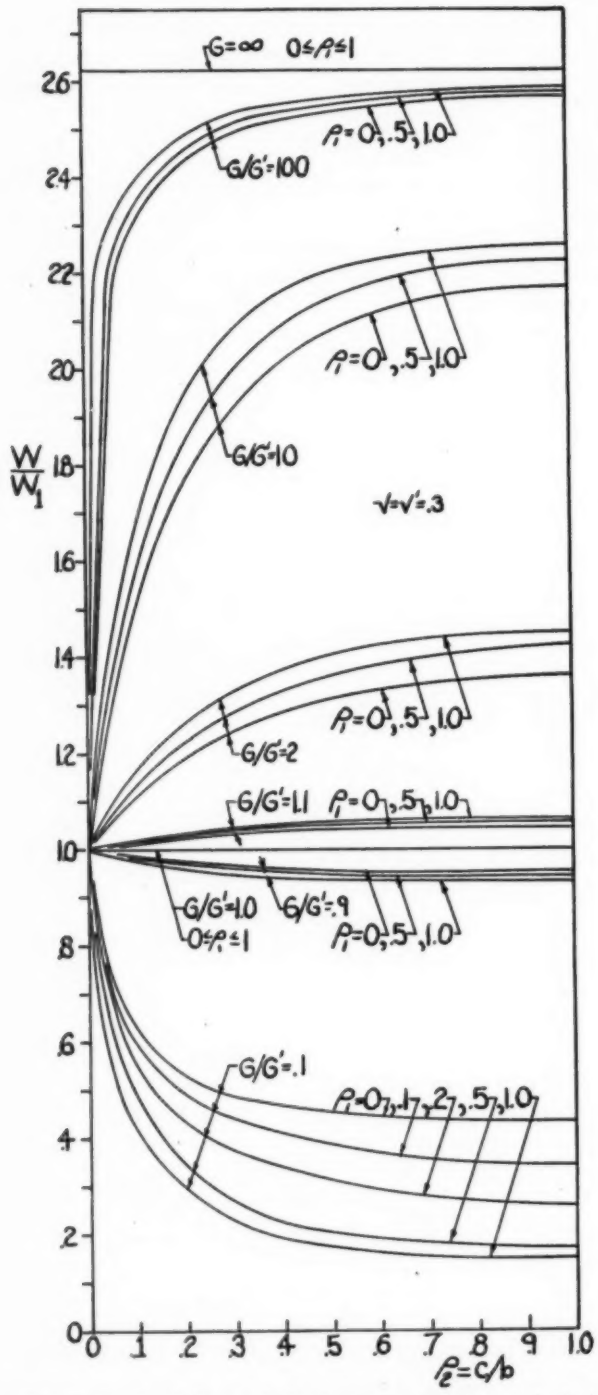


FIG. 1. Ratio of the total elastic energy (W) for an inclusion with shape ratios ρ_1, ρ_2 to the total elastic energy (W_1) for an elliptic disk as a function of ρ_1, ρ_2, G/G'.

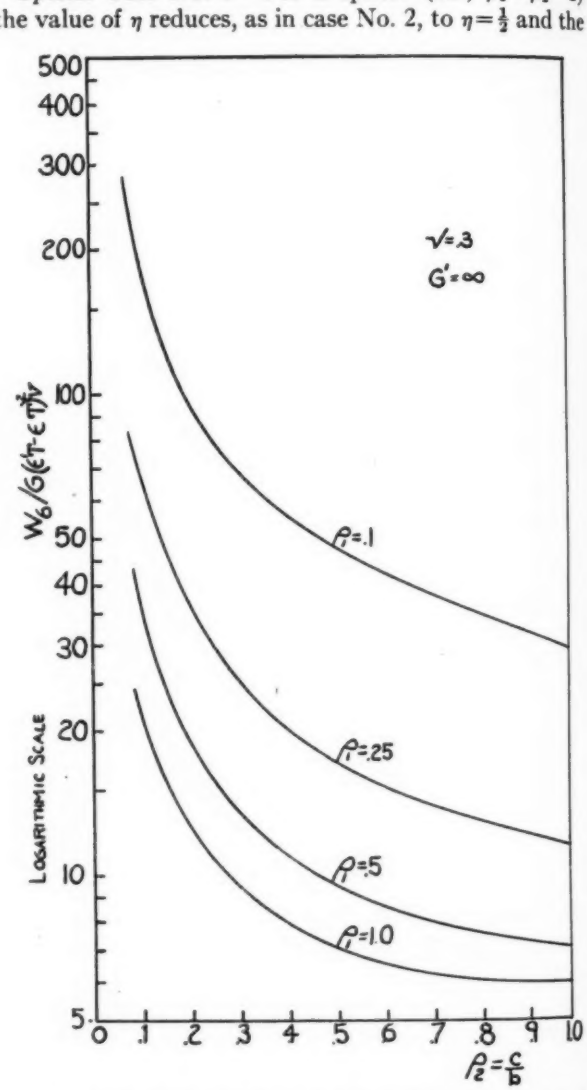


FIG. 2. $W/G(\epsilon'T-\epsilon T)^2V$ as a function of ρ_1, ρ_2, for $G'=\infty$, $(\nu=\nu'=0.3)$.

[13] For this solution see A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover Publications, New York, 1944).

function W becomes

$$W=\frac{3G'(\epsilon'T-\epsilon T)^2V}{[G/2G']+[(1-2\nu')/(1+\nu')]}, \quad (8.05)$$

which is easily verified from Lame's solution[14] for pressure applied to a concentric spherical hole in a sphere.

It is worthy of note that those cases for which $\eta=\frac{1}{2}$ (i.e., cases 2 and 5) correspond to the cases where the displacement is the gradient of a potential function and the solution may be found from the integration of a distribution of centers of dilatation (see reference 1). Further the dilatation in the interior is related to the sum of the principal stresses in the interior by $\theta'=-4G\nabla\cdot\mathbf{u}'$, while the dilatation vanishes in the exterior region. We now consider an inclusion of arbitrary shape subject to a distribution of temperature change T and having a shear modulus G, Poisson's ratio ν', and coefficient of expansion ϵ'. The exterior, to which the inclusion is bonded, has the same shear modulus G, a different Poisson's ratio ν and, for convenience, no temperature change. It is readily verified from potential theory that

$$\mathbf{u}=-[(1+\nu')/4\pi(1-\nu')]\nabla\int\int\int_{V'}[\epsilon'T/r]dV \quad (8.06)$$

(where r is the distance from the element of volume to the point where the potential is being measured) satisfies the equation of equilibrium (1.03) for both the interior and the exterior, that the tractions are continuous on the surface of the interior volume, and that the displacement and stress vanish at infinity. Hence Eq. (8.06) is the solution to the problem. Further the result of special case No. 2 ($G=G'$ with $\epsilon T=0$) that

$$W_1=2G(\epsilon'T)^2V(1+\nu')/(1-\nu') \quad (8.07)$$

is also true for an arbitrary shape if the shear moduli of the interior and exterior are equal.

From Eq. (8.01) W is seen to be a linear function of G', V, $(\epsilon'T-\epsilon T)^2$. For dependence on ρ_1, ρ_2, G/G' we consider the dimensionless ratio of the energy of a given shape (ρ_1, ρ_2) to that of the flat elliptic disk (i.e., $\rho_1=0$) for inclusions having the same shear moduli, Poisson's ratios, volume and value of $(\epsilon'T-\epsilon T)$. The ratio then becomes

$$\frac{W}{W_1}=\frac{(\frac{1}{2})+[(1-2\nu')/(1+\nu')]}{[\eta G'/G]+[(1-2\nu')/(1+\nu)]}, \quad (8.08)$$

where η is given by Eq. (7.05). The dependence of this function (for $\nu=\nu'=0.3$) upon the variables ρ_1, ρ_2, G/G' is shown in Fig. 1, where it can be seen: (1) that W/W_1 increases monotonically with respect to increasing G/G'; (2) for values of $G/G'>1$ the function W/W_1 increases monotonically with respect to increasing ρ_1 and ρ_2; (3) for values of $G/G'<1$, the function W/W_1, decreases monotonically with respect to increasing ρ_1, or ρ_2; (4) for $G/G'=\infty$ (i.e., $G=\infty$) the value W/W' has an upper limit which is

$$W/W_1=\tfrac{3}{2}(1-\nu')/(1-2\nu'). \quad (8.09)$$

For $G/G'=0$ (i.e., $G=0$) the function W/W' has a lower limit which is a function of ρ_1, ρ_2, ν, ν'. W, how-

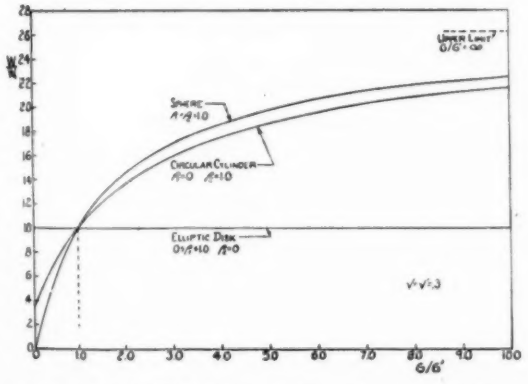


FIG. 3. Ratio W/W', for a sphere, circular cylinder and an elliptic disk as a function of G/G'.

[14] See reference 13, p. 142.

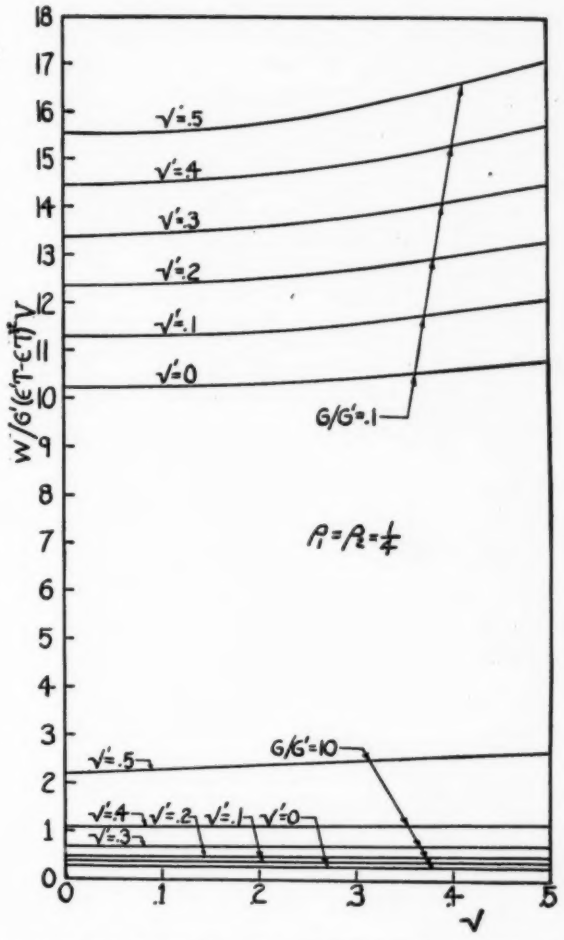


FIG. 4. $W/G'(\epsilon'T-\epsilon T)^2V$ as a function of ν, ν' for an ellipsoidal inclusion with shape ratios $\rho_1=\rho_2=\frac{1}{4}$ and $G/G'=0.1$, 10.

ever, vanishes for all shape ratios if the values are approached in the order 1: $G/G'=0$; 2: remaining variables i.e., ρ_1, ρ_2, ν, ν'. In cases 1, 3, 4, this order was not preserved and the resulting values of W_1, W_3, W_4 are not zero for $G/G'=0$. Thus we see that the limits of the functions W, W_1, W/W_1 are not uniform at $G/G'=0$, for values of $\rho_1=0$ or $\rho_2=0$.

Values of the functions along $\rho_2=1$ (prolate spheroid) agree with those obtainable from Edwards,[5] results. The same non-uniform limit condition at $G/G'=0$, $\rho_1=0$ is found in his formulas.

Formula (8.01) becomes indeterminate in its present form if $G'=\infty$ (i.e., rigid interior), but it may be put in an alternate form by multiplying numerator and denominator by G/G' and then setting $G'=\infty$. The resulting formula may be written in the dimensionless form,

$$W_0/G(\epsilon'T-\epsilon T)^2V=3/\eta. \tag{8.10}$$

This function is given (logarithmically) as a function of ρ_1, ρ_2 in Fig. 2 where we note the non-existence of its limit as $\rho_1\to 0$ or $\rho_2\to 0$ and that it decreases monotonically with respect to increasing ρ_1 or ρ_2.

Figure 3 indicates the value of the ratio W/W' for the limiting shapes; sphere, circular cylinder, and elliptic disk as they depend upon the ratio G/G' for $(\nu=\nu'=.3)$. It is to be noted that if the shear modulus of the inclusion (G') is greater than that of the matrix (G) the shape associated with the minimum total elastic energy is the sphere. If the shear modulus of the inclusion is less than that of the matrix the shape associated with the minimum total elastic energy is the elliptic disk. For equal shear modulii the energy is independent of both the shape and the interior Poisson's ratio and is given by Eq. (8.07).

Figure 4 indicates the dependence of the quantity $W/G'(\epsilon'T-\epsilon T)^2V$ on ν and ν' for a shape $\rho_1=\rho_2=\frac{1}{4}$ and two values of G/G'.

9. ACKNOWLEDGMENT

I wish to acknowledge, with thanks, the invaluable advice of Professor R. D. Mindlin in connection with the solution of this problem. I wish also to thank Professors M. A. Sadowsky and E. Sternberg for making available to me their detailed computations for the ellipsoidal cavity problem.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

The One-Sided Green's Function

KENNETH S. MILLER
New York University, New York, New York
(Received March 7, 1951)

Let L be a linear differential operator of the nth order whose coefficients $p_i(x)$ are continuous in a semi-infinite interval I: $[a, \infty)$. A function $H(x, \xi)$ is said to be a one-sided Green's function for the operator L if it satisfies the four conditions: (1) H is continuous and its first n derivatives with respect to x are continuous in I. (2) $H_\alpha(\xi, \xi)=0$ for $\alpha=0, 1, \cdots, n-2$. (3) $H_{n-1}(\xi, \xi)=1/p_0(\xi)$. (4) $LH=0$. (The subscript on the H refers to partial differentiation with respect to the first argument, and $p_0(x)$ is the coefficient of d^n/dx^n in the expression for L.) It is shown that H is unique and if $u(x)=\int_a^x H(x, \xi)f(\xi)d\xi$, then $Lu=f(x)$ and $u^{(\alpha)}(a)=0$, $\alpha=0, 1, \cdots, n-1$. Furthermore, if H is given, a fundamental system of solutions of $Lu=0$ can be written down explicitly in terms of H and its derivatives evaluated at the end point a. The converse problem is trivial. Other properties of H are also considered, for example, its relation to the impulsive response of a network.

1. INTRODUCTION

THE classical Green's function is one of the most powerful tools available for the investigation of linear differential equations, and hence all physical systems which are governed by a linear differential system. Roughly speaking, if L is a linear differential operator with which are associated linear two-point boundary conditions in the interval $[a, b]$, then a function $G(x, \xi)$ called the *Green's function* of the system can be defined.[1] This function has the property, among others, that if $u(x)=\int_a^b G(x, \xi)f(\xi)d\xi$, then $Lu(x)=f(x)$.

Now with many problems in engineering (particularly in network theory) we must associate *one-point* boundary conditions. Furthermore, the solutions is generally of interest not only in a finite interval, but, say, for all time with $t\geqq 0$. For problems of this nature, a more convenient function to use than the *classical Green's function* $G(x, \xi)$ is a function $H(x, \xi)$ which we shall call the *one-sided Green's function*. [$H(x, \xi)$ is formally defined in Definition 1 below.] It will be seen that $H(x, \xi)$ is, in general, an easier function to deal with (it having no discontinuities) than the corresponding classical Green's function $G(x, \xi)$. Also, $H(x, \xi)$ is defined only with respect to the differential *operator* L, while to define $G(x, \xi)$, the differential *system* must be taken into account.

[1] E. L. Ince, *Ordinary Differential Equations* (Dover Publications, New York, 1944), p. 254.

The motivation underlying the formulation of the one-sided Green's function rests in the method of variation of parameters used in solving nonhomogeneous differential equations. (See, e.g., reference 1, p. 122.) The formal definition of $H(x, \zeta)$ reads as follows:

Definition 1. Let

$$L=p_0(x)(d^n/dx^n)+p_1(x)(d^{n-1}/dx^{n-1})+\cdots+p_n(x)\cdot \quad (1)$$

be a linear differential operator whose coefficients $p_i(x)$ are real continuous functions in the semi-infinite interval I: $[a, \infty)$, and $p_0(x)>0$ in any closed bounded subset of I. We define a one-sided Green's function of the linear differential operator L as any real function $H(x, \zeta)$ of the two real variables x and ζ such that

(1) $H(x, \zeta)$ is continuous in both variables simultaneously and the first n derivatives with respect to x are continuous in I;

(2) $\partial^\alpha H(x, \zeta)/\partial x^\alpha|_{x=\zeta}=0$ for $\alpha=0, 1, \cdots, n-2$;

(3) $\partial^{n-1}H(x, \zeta)/\partial x^{n-1}|_{x=\zeta}=1/p_0(\zeta)$;

(4) $LH(x, \zeta)=0$.

2. PROPERTIES OF $H(x,\zeta)$

A. Construction of a One-Sided Green's Function

Before proving any theorems about H we shall actually construct a one-sided Green's function in terms of a fundamental set of solutions of the linear differential operator L. Let L be as in Eq. (1) and let $\{\phi_1(x), \phi_2(x), \cdots, \phi_n(x)\}$ be a fundamental set of solutions of the linear differential equation $Lu=0$. Then we have

$$H(x, \zeta)=\frac{(-1)^{n-1}}{p_0(\zeta)W(\zeta)}\begin{vmatrix} \phi_1(x) & \phi_2(x) & \cdots & \phi_n(x) \\ \phi_1(\zeta) & \phi_2(\zeta) & \cdots & \phi_n(\zeta) \\ \phi_1'(\zeta) & \phi_2'(\zeta) & \cdots & \phi_n'(\zeta) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_1^{(n-2)}(\zeta) & \phi_2^{(n-2)}(\zeta) & \cdots & \phi_n^{(n-2)}(\zeta) \end{vmatrix}, \quad (2)$$

where

$$W(\zeta)=\begin{vmatrix} \phi_1(\zeta) & \phi_2(\zeta) & \cdots & \phi_n(\zeta) \\ \phi_1'(\zeta) & \phi_2'(\zeta) & \cdots & \phi_n'(\zeta) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_1^{(n-1)}(\zeta) & \phi_2^{(n-1)}(\zeta) & \cdots & \phi_n^{(n-1)}(\zeta) \end{vmatrix} \quad (3)$$

is the Wronskian of $\{\phi_1(\zeta), \phi_2(\zeta), \cdots, \phi_n(\zeta)\}$. That this $H(x, \zeta)$ satisfies conditions (1)–(4) of Definition 1 is readily verified by a direct calculation.

B. Uniqueness of H(x, ζ)

Theorem 1. The Green's function $H(x, \zeta)$ constructed for the linear differential operator is unique.

Proof. From Eq. (2) we may write

$$H(x, \zeta)=\sum_{i=1}^{n} a_i(\zeta)\phi_i(x),$$

where the $\phi_i(x)$ are a fundamental system for L. Introduce the notation

$$H_\alpha(x, \zeta)=\partial^\alpha H(x, \zeta)/\partial x^\alpha.$$

Then, we have

$$H_\alpha(\zeta, \zeta)=\sum_{i=1}^{n} a_i(\zeta)\phi_i^{(\alpha)}(\zeta)=0, \quad \alpha=0, 1, \cdots, n-2,$$

$$H_{n-1}(\zeta, \zeta)=\sum_{i=1}^{n} a_i(\zeta)\phi_i^{(n-1)}(\zeta)=1/p_0(\zeta).$$

Suppose $H(x, \zeta)$ is not unique. Let $K(x, \zeta)$ be another one-sided Green's function for L. Let

$$J(x, \zeta)=H(x, \zeta)-K(x, \zeta)=\sum_{i=1}^{n} b_i(\zeta)\phi_i(x).$$

Then we have

$$\sum_{i=1}^{n} b_i(\zeta)\phi_i^{(\alpha)}(\zeta)=0, \quad \alpha=0, 1, \cdots, n-1.$$

Since the determinant of the $b_i(\zeta)$ is the Wronskian of the $\phi_i(\zeta)$, we have $b_i(\zeta)\equiv 0$, $i=1, 2, \cdots, n$, all ζ. Hence, we have $J(x, \zeta)\equiv 0$, and $H(x, \zeta)$ is unique.

C. Solution of $Lu=f$

Theorem 2. Let L be a linear differential operator as in Eq. (1). Let $H(x, \zeta)$ be the one-sided Green's function for L, and let $f(x)$ be any real continuous function in I. Then, if

$$u(x)=\int_a^x H(x, \zeta)f(\zeta)d\zeta,$$

it follows that

$$Lu(x)=f(x).$$

Proof. We may write

$$u'(x)=H(x,x)f(x)+\int_a^x H_1(x,\zeta)f(\zeta)d\zeta$$

$$=\int_a^x H_1(x,\zeta)f(\zeta)d\zeta,$$

since $H(x,x)\equiv 0$ by definition of H. Similarly, we have

$$u^{(k)}(x)=\int_a^x H_k(x,\zeta)f(\zeta)d\zeta, \quad k=0,1,\cdots,n-1,$$

and

$$u^{(n)}(x)=H_{n-1}(x,x)f(x)+\int_a^x H_n(x,\zeta)f(\zeta)d\zeta$$

$$=[f(x)/p_0(x)]+\int_a^x H_n(x,\zeta)f(\zeta)d\zeta.$$

If we multiply $u^{(k)}$ by $p_{n-k}(x)$ and add, we obtain

$$Lu=f(x)+\int_a^x LH(x,\zeta)f(\zeta)d\zeta=f(x).$$

D. Equivalence of $H(x,\zeta)$ and a Fundamental System

In this section we shall show that if the one-sided Green's function $H(x,\zeta)$ of a linear differential operator L is known, then n linearly independent solutions of $Lu=0$ can be explicitly written down.[2] [In view of our construction in Sec. 2. A the converse problem is trivial, that is, to find $H(x,\zeta)$ given a fundamental system $\{\phi_1(x), \phi_2(x), \cdots, \phi_n(x)\}$.] In order to expedite the proof of Theorem 3 we state a lemma[3] and sketch its proof.

Lemma 1. Let $\{\phi_1(x), \phi_2(x), \cdots, \phi_n(x)\}$ be a fundamental set of solutions of $Lu=0$ [see Eq. (1)] with Wronskian $W(x)$. Then

$$v_i(x)=[1/p_0(x)]\partial \log W(x)/\partial\phi_i^{(n-1)}, \quad i=1,2,\cdots,n,$$

form a fundamental set of solutions of the adjoint equation $L^+v=0$.[4]

Proof. Form the linear expression

$$M_iu=\frac{1}{W}\sum_{\alpha=0}^{n-1}\frac{\partial W}{\partial\phi_i}u^{(\alpha)}.$$

Then it follows that $M_i\phi_j=\delta_{ij}$, where δ_{ij} is the Kronecker delta. Hence, if u is any linear combination of the $\phi_i(x)$, such as

$$u(x)=C_1\phi_1(x)+C_2\phi_2(x)+\cdots+C_n\phi_n(x),$$

we have $M_iu=C_i$, or

$$dM_iu/dx=0.$$

Since dM_iu/dx is of the nth order in the derivatives of u and has the same fundamental system as $Lu=0$, dM_iu/dx must differ from Lu by at most a nonzero multiplicative constant, say, $v_i(x)$,

$$dM_iu/dx=v_iLu.$$

The factor v_i can be determined by equating coefficients of $u^{(n)}$ in dM_iu/dx and Lu. Doing this, we obtain

$$v_i=(1/p_0)(1/W)\partial W/\partial\phi_i^{(n-1)}=(1/p_0)\partial \log W/\partial\phi_i^{(n-1)}.$$

Since v_iLudx is an exact differential, $v_i(x)$ must satisfy the adjoint equation $L^+u=0$.

Theorem 3. Hypothesis. Let

$$L=p_0(x)(d^n/dx^n)+p_1(x)(d^{n-1}/dx^{n-1})+\cdots+p_n(x).$$

be a linear differential operator with coefficients $p_i(x)\epsilon C^{n-i}$ in the semi-infinite interval I: $[a,\infty)$. Let $p_0(x)>0$ in any closed bounded subset of I. Let $H(x,\zeta)$ be the one-sided Green's function of the operator L.

Conclusion. The functions $\{g_\alpha(x)\}$, $\alpha=1,2,\cdots,n$ form a fundamental system of solutions for $Lu=0$, where

$$g_\alpha(x)=[1/p_0(x)]\partial \log\omega(x)/\partial\psi_\alpha^{(n-1)}, \quad \alpha=1,2,\cdots,n,$$

$$\psi_\beta(\zeta)=\partial^{\beta-1}H(x,\zeta)/\partial x^{\beta-1}\big|_{x=a}, \quad \beta=1,2,\cdots,n,$$

and $\omega(x)$ is the Wronskian of the $\psi_\beta(x)$.

Proof. From Eq. (2) we may write

$$H(x,\zeta)=\sum_{i=1}^n \phi_i(x)\phi_i^*(\zeta),^5$$

where the $\phi_i(x)$ are n linearly independent solutions of $Lu=0$ and the $\phi_i^*(\zeta)$ are n linearly independent solutions of $L^+u=0$. We construct the following derivatives:

$$\psi_{\alpha+1}(\zeta)\equiv\frac{\partial^\alpha}{\partial x^\alpha}H(x,\zeta)\bigg|_{x=a}=\sum_{i=1}^n \phi_i^{(\alpha)}(a)\phi_i^*(\zeta) \quad (4)$$

for $\alpha=0,1,\cdots,n-1$. Since the $\psi_\beta(\zeta)$ are linear combinations of the $\phi_i^*(\zeta)$ (with constant coefficients), they are solutions of $L^+u=0$. We now show that the $\psi_\beta(\zeta)$ are linearly independent. Write Eq. (4) in the matrix form $\psi=W\phi^*$, where ψ is the column vector of $\psi_{\alpha+1}(\zeta)$, $\alpha=0,1,\cdots,n-1$, W is the square matrix $\|\phi_i^{(\alpha)}(a)\|$, $i=1,2,\cdots,n$, $\alpha=0,1,\cdots,n-1$, and ϕ^* is the column vector of the $\phi_i(\zeta)$, $i=1,2,\cdots,n$. [Since W is the matrix of the Wronskian of the $\phi_i(x)$ evaluated at $x=a$, W is nonsingular.] If the ψ_β were linearly dependent, there would exist a constant vector $c\neq 0$ such that

[2] An analogous problem (among others) is solved for the classical Green's function in a forthcoming paper by K. S. Miller and M. M. Schiffer entitled "On the Green's functions of ordinary differential systems."

[3] G. Darboux, *Leçons sur la théorie générale des surfaces* (Gauthier-Villars et Fils, Paris, 1889), Vol. II, p. 102.

[4] L^+ is the formal or lagrange adjoint of L. See reference 1, p. 123.

[5] Reference 1, p. 256.

$c\psi \equiv 0$. This implies

$$0 \equiv cW\phi^*. \tag{5}$$

It is also the case that $cW \not\equiv 0$, since W is nonsingular; and hence Eq. (5) implies that the ϕ_i^* are linearly dependent—a contradiction.

If we let

$$g_\alpha(x) = [1/p_0(x)]\partial \log\omega(x)/\partial\psi_\alpha^{(n-1)}, \quad \alpha = 1, 2, \cdots, n, \tag{6}$$

where $\omega(x)$ is the Wronskian of the $\psi_\alpha(x)$, then $g_i(x)$, $i=1, 2, \cdots, n$ are n linearly independent solutions of $Lu=0$. (See Lemma 1.)

E. The Wronskian

The Wronskian $W(x)$ of the n linearly independent functions $g_\alpha(x)$ of the above theorem can be written down as a determinant. However, we can obtain a simple expression for $W(x)$ in terms of the Green's function. If we compute $H_n(\zeta, \zeta)$ from Eq. (2) and use the fact that

$$p_0 g^{(n)} = -p_1 g^{(n-1)} - p_2 g^{(n-2)} - \cdots - p_n g,$$

where g is any of the n linearly independent solutions of $Lu=0$, we have

$$H_n(\zeta, \zeta) = -p_1(\zeta)/p_0^2(\zeta)$$

and

$$H_n(\zeta, \zeta)/H_{n-1}(\zeta, \zeta) = -p_1(\zeta)/p_0(\zeta).$$

Now $W(x)$ satisfies $p_0W' + p_1W = 0$. Hence, if we let

$$k(\zeta) = H_n(\zeta, \zeta)/H_{n-1}(\zeta, \zeta),$$

$$W(x) = W_0 \exp\left[\int_a^x k(\zeta)d\zeta\right],$$

where W_0 is a nonzero constant.

3. DISCUSSION OF $H(x, \zeta)$

If we have the relation

$$u(x) = \int_a^x H(x, \zeta)f(\zeta)d\zeta, \tag{7}$$

we note that $u^{(\alpha)}(a) = 0$, $\alpha = 0, 1, \cdots, n-1$. That is, the boundary conditions on $u(x)$ are on only one point of the interval—hence, the term "one-sided." The classical Green's function $G(x, \zeta)$ for the totally homogeneous incompatible linear differential system

$$Lu = 0,$$

$$U_\alpha(u) = \sum_{i=1}^{n-1} A_{\alpha i}u^{(i-1)}(a) + \sum_{i=1}^{n-1} B_{\alpha i}u^{(i-1)}(b) = 0,$$

$\alpha = 1, 2, \cdots, n$ (for some closed finite interval $[a, b]$), has the property that

$$U_\alpha[G(x, \zeta)] = 0, \quad \alpha = 1, 2, \cdots, n.$$

In the case of the one-sided Green's function, $H(x, \zeta)$ does *not* satisfy the boundary conditions $H_\alpha(a, \zeta) = 0$. Also, $H(x, \zeta)$ has no discontinuities, while $G(x, \zeta)$ has a discontinuity in its $(n-1)$st derivative at the point $x = \zeta$.

If we attempt to impose more general boundary conditions on the system, we must use two-point boundary conditions. In this case the Volterra-type integral equation involving $H(x, \zeta)$ must be modified to a Fredholm type, and we shall find that $H(x, \zeta)$ has to be modified to the classical two-sided Green's function. However, as mentioned before, many engineering problems have only one-sided boundary conditions;[6] and hence $H(x, \zeta)$ is admirably suited to deal with such problems.

As another example (besides the above quoted reference) we mention the connection between the one-sided Green's function and the impulsive response of a linear network.[7] Suppose we have a linear network N with system function $Y(p)$, p being the heaviside operator d/dt. Then the output of the system, $y(t)$, is related to the input $x(t)$ of N by the operational equation

$$y(t) = Y(p)x(t).$$

Assume, as is often the case, that the inverse of Y is a linear differential operator L, then we have

$$Ly(t) = x(t).$$

If the system is assumed to be unexcited, and $x(t)$ is not applied until $t=0$, we may write, in our mathematical notation,

$$y(t) = \int_0^t H(t, \zeta)x(\zeta)d\zeta,$$

where $H(t, \zeta)$ is the one-sided Green's function for the linear differential operator L.

Now the impulsive response $W(t)$ of the network N is defined as the response when a delta-function is applied at the input. If we take the function $\delta(t-\tau)$ as the input, that is, apply a delta-function at time $t=\tau$, then we obtain

$$x(t) = \delta(t-\tau)$$

and

$$W(t) = \int_0^t H(t, \zeta)\delta(\zeta-\tau)d\zeta = H(t, \tau).$$

We sometimes write $W(t, \tau)$ instead of $W(t)$ to indicate the dependence of W on the parameter τ.

---

[6] See, for example, K. S. Miller and R. J. Schwarz, J. Appl. Phys. **21**, 290–294 (1950). However, precisely because of its one-sidedness, $H(x, \zeta)$ does not readily lend itself to many theoretical problems, for example, determining the conditions under which Eq. (7) represents a self-adjoint integral operator.

[7] L. A. Zadeh, J. Appl. Phys. **21**, 642–645 (1950).

The Impedance of an Antenna over a Large Circular Screen

JAMES E. STORER*
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received February 20, 1951)

In the performance of antenna measurements the antenna is usually placed over a large conducting screen. The assumption is then made that the measurements are in close agreement with those that would be obtained if the screen were of infinite size. Recent experimental work, however, indicates that the finiteness of the ground screen may influence the measurements to a considerable extent.

It is therefore desirable to solve this problem theoretically in order to determine the extent of the change introduced by the use of a finite screen. This paper treats the problem of an antenna erected vertically in the center of a circular screen. An integral equation is obtained for the electric field in the plane of the screen, outside the screen. This equation is solved approximately by the use of a variational principle, and a formula is obtained for the change of antenna impedance as a function of the diameter of the screen. The formula may be used to give a simple and accurate estimate of the changes introduced by the use of a finite screen.

I. INTRODUCTION

IN most measurements made on thin wire antennas, it is customary to place the antenna over a large conducting screen in the manner indicated in Fig. 1. If this screen is large in comparison with the wavelength being used, the assumption is generally made that the experimental results are in close agreement with those that would be obtained if the screen were of infinite size. Recently, Meier and Summers[1] measured experimentally some effects produced by using a finite screen. Their results indicated that measurements taken over a finite screen may differ appreciably from those that would be taken using an infinite screen.

It would therefore be desirable to solve this problem theoretically and determine the effect of using a finite screen. To the best of the author's knowledge, the first person to attempt to do this was Bardeen.[2] He considered the problem of an antenna placed vertically in the center of a circular screen and obtained an integral equation for the currents on the screen. However, he did not solve this equation except for the case of a screen small in comparison with the wavelength.

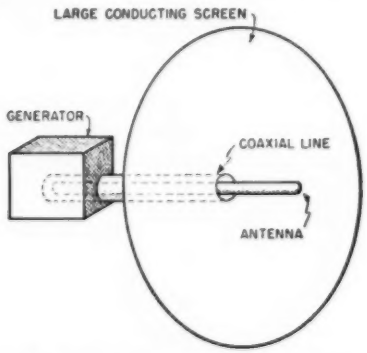


FIG. 1. Typical experimental setup for performing measurements on antennas.

Recently, Leitner and Spence[3] obtained a solution for this problem in the form of an infinite series of spheroidal functions. Unfortunately, however, this series converges very slowly for the larger radii screens. Thus, for the practical case of screens greater than 10 wavelengths in diameter, this solution is limited in its usefulness. Moreover, Leitner and Spence confine themselves to the particular case of an infinitesimally thin quarter-wavelength antenna. As will be seen, this is a very serious restriction, since the effects of a finite screen change markedly for different lengths of antenna.

This paper will consider the same problem, i.e., an antenna erected vertically in the center of a circular screen. However, there will be no restriction placed on the length of the antenna. The problem is formulated in terms of an integral equation for the radial component of the electric field in the plane of the screen. Then, for the useful case of large diameter screens, this integral equation is solved approximately by means of a variational method. A simple and easily evaluated expression for the change of the antenna impedance as a function of the screen diameter is obtained. A discussion of the change of the fields will appear in a later paper.

The formula for the change of the impedance is a useful one. It gives a very simple criterion for determining whether a ground screen is large enough to be considered, for practical purposes, an infinite screen.

II. FORMULATION OF THE PROBLEM

The physical problem will be simplified by assuming that the screen is of vanishing thickness and perfectly conducting. Since this paper is primarily concerned with large screens, the apparatus to the rear of the screen driving the antenna may be ignored.

A cylindrical coordinate system (r, ϕ, z) will be used, oriented so that the z axis coincides with the center of the antenna and the screen lies in the $z=0$ plane (see

* The research reported in this document was carried out under a fellowship awarded the author by the AEC. Publication of the report has been made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR), the Signal Corps of the U. S. Army, and the U. S. Air Force, under ONR Contract N5-ori-76, T. O. 1.

[1] A. S. Meier and W. P. Summers, Proc. Inst. Radio Engrs. **37**, 609–616 (1949).

[2] J. Bardeen, Phys. Rev. **36**, 1482–1488 (1930).

[3] A. Leitner and R. D. Spence, J. Appl. Phys. **21**, 1001–1006 (1950).

Fig. 2). The height of the antenna will be represented by h and the diameter of the screen by d. The total current flowing at a point z on the antenna will be denoted by $I(z)$. Since this paper is concerned only with fields at far distances from the antenna, the antenna may—and will be—represented by a filament of current, $I(z)$, concentrated on the z axis.

Obviously, a circular symmetry exists and there is no dependence of the fields on the ϕ-coordinate. The only electromagnetic-field components present are, therefore, E_r, E_z, and H_ϕ. When a harmonic time dependence of the form $e^{-i\omega t}$ is used, the maxwell equations for the fields become

$$E_r=\zeta_0\partial H_\phi/ik\partial z,$$
$$E_z=-\zeta_0\partial rH_\phi/ikr\partial r, \tag{1}$$

$$(\partial^2H_\phi/\partial z^2)+(\partial^2H_\phi/\partial r^2)+(\partial H_\phi/r\partial r)-(H_\phi/r^2)+k^2H_\phi=0, \tag{2}$$

where

$$k=\omega(\mu_0\epsilon_0)^{\frac{1}{2}},$$
$$\zeta_0=(\mu_0/\epsilon_0)^{\frac{1}{2}}=120\pi \text{ ohms}.$$

The boundary conditions associated with these equations are (1) that E_r must vanish on the screen, (2) that the fields present are "radiating" fields, and (3) that

$$\lim_{r\to 0} H_\phi=\frac{I(z)}{2\pi r}, \quad 0<z\leq h. \tag{3}$$

This last condition is simply the mathematical statement of the fact that the current $I(z)$ is exciting the fields.

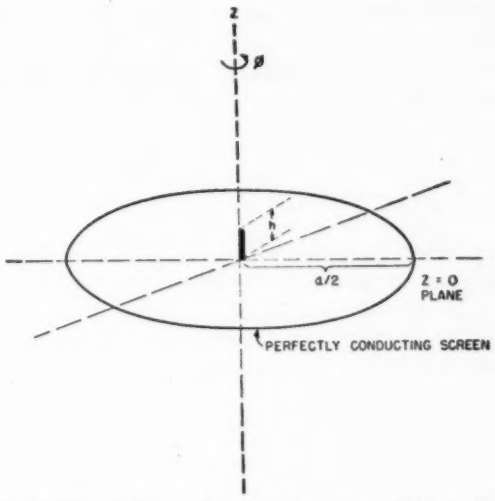


FIG. 2. Orientation of the cylindrical coordinate system.

Since the screen is assumed to be large, the fields present do not differ greatly from those of an infinite screen. Hence, it is logical to set

$$H_\phi=H_\phi^\infty+\bar{H}_\phi,$$

where H_ϕ^∞ is the magnetic field when an infinite screen is present, $\bar{H}_\phi$ is the change of the field when a finite screen is present.

H_ϕ^∞ may be obtained immediately by using a Helmholtz integral together with the theory of images. Thus, one obtains

$$H_\phi^\infty(r,z)=\begin{cases}-\dfrac{1}{4\pi}\dfrac{\partial}{\partial r}\underset{\text{antenna and image}}{\iiint}\dfrac{\exp\{ik[(z-z')^2+r^2+r'^2-2rr'\cos(\phi-\phi')]^{\frac{1}{2}}\}}{[(z-z')^2+r^2+r'^2-2rr'\cos(\phi-\phi')]^{\frac{1}{2}}}j(r',z')dv' & z>0\\ 0 & z<0\end{cases},$$

where $j(r,z)$ is the current density on the antenna.

Since the antenna is being represented by its total current, $I(z)$, concentrated on the z axis, this integral becomes

$$H_\phi^\infty(r,z)=\begin{cases}\underbrace{-\dfrac{1}{4\pi}\dfrac{\partial}{\partial r}\displaystyle\int_0^h\frac{\exp\{ik[(z-z')^2+r^2]^{\frac{1}{2}}\}}{[(z-z')^2+r^2]^{\frac{1}{2}}}I(z')dz'}_{\text{antenna}} \underbrace{-\dfrac{1}{4\pi}\dfrac{\partial}{\partial r}\displaystyle\int_0^h\frac{\exp\{ik[(z+z')^2+r^2]^{\frac{1}{2}}\}}{[(z+z')^2+r^2]^{\frac{1}{2}}}I(z')dz'}_{\text{image}} & z>0\\ 0 & z<0\end{cases}.$$

From the manner in which it was constructed, it is apparent that H_ϕ^∞ must satisfy the wave equation (2) and the condition

$$(E_r^\infty)_{z=0}=(\zeta_0/ik)(\partial H_\phi^\infty/\partial z)_{z=0}=0.$$

Now, since both H_ϕ and H_ϕ^∞ satisfy the wave equation (2), it follows that $\bar{H}_\phi$ must also satisfy this equation, i.e.,

$$[(\partial^2/\partial z^2)+(\partial^2/\partial r^2)+(\partial/r\partial r)-(1/r^2)+k^2]\bar{H}_\phi=0. \tag{4}$$

$\bar{H}_\phi$ is continuous everywhere except across the $z=0$ plane, where it is discontinuous because of the presence of the screen. Let $J_1(x)$ denote a bessel function of the first order and kind. It can be shown[4] that any radiating solution to the wave equation (4), discontinuous across the $z=0$ plane, may be repre-

[4] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), pp. 369–371.

sented in the form

$$\bar{H}_\phi(r,z)=\begin{cases}\displaystyle\int_0^\infty J_1(\lambda r)\exp[i(k^2-\lambda^2)^{\frac{1}{2}}z]f^+(\lambda)\lambda d\lambda, & z>0\\ \displaystyle\int_0^\infty J_1(\lambda r)\exp[-i(k^2-\lambda^2)^{\frac{1}{2}}z]f^-(\lambda)\lambda d\lambda, & z<0\end{cases}.$$

The positive imaginary character of the exponentials insures that these integrals represent radiating functions. The $(k^2-\lambda^2)^{\frac{1}{2}}$ is defined to be equal to $i(\lambda^2-k^2)^{\frac{1}{2}}$ for $\lambda>k$ to secure convergence of the integrals. The boundary conditions at $z=0$ must be used to determine the functions $f^+(\lambda)$ and $f^-(\lambda)$.

The first boundary condition that will be imposed is that E_r must vanish on the screen. Remembering that $(E_r^\infty)_{z=0}=0$, $(E_r)_{z=0}$ is given by

$$(E_r)_{z=0}=(E_r^\infty+\bar{E}_r)_{z=0}=(\bar{E}_r)_{z=0}=(\zeta_0/ik)(\partial\bar{H}_\phi/\partial z)_{z=0},$$

$$(E_r)_{z=0}=\begin{cases}0, & 0<r<d/2 \quad \text{(on screen)}\\ \xi(r), & r>d/2\end{cases}.$$

The function $\xi(r)$ thus has the interpretation of being the radial component of the electric field in the plane of the screen and outside the screen. Inserting the integral represetations of $\bar{H}_\phi$, one obtains the following equations:

$$\frac{\zeta_0}{k}\int_0^\infty J_1(\lambda r)[k^2-\lambda^2]^{\frac{1}{2}}f^+(\lambda)\lambda d\lambda=\begin{cases}0, & 0<r<d/2\\ \xi(r), & r>d/2\end{cases}$$

$$-\frac{\zeta_0}{k}\int_0^\infty J_1(\lambda r)[k^2-\lambda^2]^{\frac{1}{2}}f^-(\lambda)\lambda d\lambda=\begin{cases}0, & 0<r<d/2\\ \xi(r), & r>d/2\end{cases}. \quad (5)$$

Equations (5) may be solved for $f^+(\lambda)$ and $f^-(\lambda)$ by means of a fourier-bessel transform pair. For functions of the first order and first kind this transform pair is as follows:

if

$$A(r)=\int_0^\infty J_1(\lambda r)B(\lambda)\lambda d\lambda,$$

then

$$B(\lambda)=\int_0^\infty J_1(\lambda r)A(r)rdr.$$

Using these to solve Eqs. (5) for $f^+(\lambda)$ and $f^-(\lambda)$, one obtains

$$f^+(\lambda)=(k/\zeta_0[k^2-\lambda^2]^{\frac{1}{2}})\int_{d/2}^\infty J_1(\lambda r)\xi(r)rdr,$$

$$f^-(\lambda)=-(k/\zeta_0[k^2-\lambda^2]^{\frac{1}{2}})\int_{d/2}^\infty J_1(\lambda r)\xi(r)rdr.$$

When these are inserted, $\bar{H}_\phi$ becomes

$$\bar{H}_\phi=\begin{cases}(k/\zeta_0)\displaystyle\int_0^\infty J_1(\lambda r)\exp(i[k^2-\lambda^2]^{\frac{1}{2}}z)\times\left\{\int_{d/2}^\infty J_1(\lambda r')\xi(r')r'dr'\right\}\times(\lambda/[k^2-\lambda^2]^{\frac{1}{2}})d\lambda, & z>0\\ -(k/\zeta_0)\displaystyle\int_0^\infty J_1(\lambda r)\exp(-i[k^2-\lambda^2]^{\frac{1}{2}}z)\times\left\{\int_{d/2}^\infty J_1(\lambda r')\xi(r')r'dr'\right\}\times(\lambda/[k^2-\lambda^2]^{\frac{1}{2}})d\lambda, & z<0\end{cases}.$$

Since $\xi(r)$ represents a radial component of the electric field at $z=0$, it vanishes for large r as $(1/r^2)e^{ikr}$. Hence, the above integrals are absolutely convergent, and the order of integration may be interchanged. Using the notation,

$$G(z-z',r,r')=\tfrac{1}{2}i\int_0^\infty J_1(\lambda r)J_1(\lambda r')\times\exp(i[k^2-\lambda^2]^{\frac{1}{2}}|z-z'|)(\lambda/[k^2-\lambda^2]^{\frac{1}{2}})d\lambda, \quad (6)$$

the magnetic field may be written in the form

$$H_\phi=H_\phi^\infty+\bar{H}_\phi=\begin{cases}H_\phi^\infty(r,z)-(2ik/\zeta_0)\displaystyle\int_{d/2}^\infty G(z,r,r')\xi(r')r'dr', & z>0\\ (2ik/\zeta_0)\displaystyle\int_{d/2}^\infty G(z,r,r')\xi(r')r'dr', & z<0\end{cases}. \quad (7)$$

$G(z-z',r,r')$ may readily be shown to be the radiating Green's function satisfying the wave equation

$$[(\partial/\partial z^2)+(\partial^2/\partial r^2)+(\partial/r\partial r)-(1/r^2)+k^2]G(z-z',r,r')=-\delta(r-r')\delta(z-z')/(r\,r')^{\frac{1}{2}}.$$

Equation (7) could have been obtained using this Green's function together with Green's theorem. The author feels that the approach used is somewhat more direct. Although of no interest so far as this paper is concerned, it may readily be shown[5] that this function is related to the free-space Green's function by

$$G(z-z',r,r')=\int_0^{2\pi}G_0(z-z',r,r',\phi-\phi')\cos(\phi-\phi')d\phi=\int_0^{2\pi}\frac{\exp\{ik[(z-z')^2+r^2+r'^2-2rr'\cos(\phi-\phi')]^{\frac{1}{2}}\}}{4\pi[(z-z')^2+r^2+r'^2-2rr'\cos(\phi-\phi')]^{\frac{1}{2}}}\times\cos(\phi-\phi')d\phi.$$

[5] H. Levine and J. Schwinger, Phys. Rev. 75, 1430 (1949).

Equation (7) expresses the magnetic field in terms of the yet unknown function $\xi(r)$. An equation for $\xi(r)$ is obtained by noting that H_ϕ must be continuous everywhere in free space. In particular, H_ϕ must be continuous on the surface $z=0$, $r>d/2$; that is to say, in the plane of the screen outside the screen. Using Eq. (7) immediately gives

$$H_\phi^\infty(r,0)-(2ik/\zeta_0)\int_{d/2}^{\infty}G(0,r,r')\xi(r')r'dr' =(2ik/\zeta_0)\int_{d/2}^{\infty}G(0,r,r')\xi(r')r'dr', \quad r>d/2,$$

or, more concisely,

$$H_\phi^\infty(r,0)=(4ik/\zeta_0)\int_{d/2}^{\infty}G(0,r,r')\xi(r')r'dr', \quad r>d/2. \quad (8)$$

Equation (8) is valied for any diameter of screen. If $\xi(r)$ is expanded in a series of spheroidal functions, Eq. (8) may be solved exactly. This, however, leads directly back to the solution of Leitner and Spence,[3] which is useful only for screens of fairly small diameter. The advantage of formulating the problem of a finite screen in terms of an integral equation lies in the fact that it is possible to obtain a simple, approximate solution to it by means of a variational method. This solution is valid for a large ground screen.

III. VARIATIONAL FORMULATION

It is possible to obtain a variational formulation for Eq. (8) as it stands. However, it is convenient first to simplify it somewhat. Since the only solutions of interest are those for large screens, this fact will be introduced explicitly at this time by the following two assumptions:

(a) $kd\gg1$,

(b) $d^2\gg h^2$.

The necessity of making assumption (b) is apparent. Quite obviously, if the height of the antenna were of comparable size to the diameter, the screen could not be considered large.

Since in Eq. (8) $r\geq d/2$, the function $H_\phi^\infty(r,0)$ may be simplified using (a) and (b). Thus, one obtains

$$H_\phi^\infty(r,0)=-\frac{1}{2\pi}\frac{\partial}{\partial r}\int_0^h\frac{\exp(ik[r^2+z'^2]^{\frac{1}{2}})}{[r^2+z'^2]^{\frac{1}{2}}}I(z')dz' \cong-(1/2\pi)(\partial/\partial r)(e^{ikr}/r)\int_0^h I(z')dz' \cong-(ike^{ikr}/2\pi r)\int_0^h I(z)dz.$$

Using this, the integral equation becomes

$$\left\{-(\zeta_0/8\pi)\int_0^h I(z)dz\right\}e^{ikr}/r =\int_{d/2}^{\infty}G(0,r,r')\xi(r')r'dr'. \quad (9)$$

Equation (9) indicates that, although the over-all magnitude of $\xi(r)$ is dependent on $I(z)$, its functional dependence on r is not. For the purposes of the variational formulation, a parameter μ will be defined as follows:

$$\mu=\frac{\left\{-(\zeta_0/8\pi)\int_0^h I(z)dz\right\}}{\int_{d/2}^{\infty}e^{ikr}\xi(r)dr}. \quad (10)$$

If Eq. (9) is multiplied through by $r\xi(r)$, integrated with respect to r from $d/2$ to ∞, and then divided by

$$\left[\int_{d/2}^{\infty}e^{ikr}\xi(r)dr\right]^2,$$

the following expression for μ results:

$$\mu=\frac{\int_{d/2}^{\infty}\int_{d/2}^{\infty}G(0,r,r')\xi(r)\xi(r')rr'drdr'}{\left[\int_{d/2}^{\infty}e^{ikr}\xi(r)dr\right]^2}. \quad (11)$$

Equation (11) can be shown to be a variational formulation of the integral equation with μ as the stationary parameter. This is readily proved by taking the first variation of μ. Thus, using Eqs. (9) and (10) together with Eq. (11), one obtains

$$\delta\mu=\frac{\int_{d/2}^{\infty}\int_{d/2}^{\infty}G(0,r,r')\{\xi(r)\delta\xi(r')+\xi(r')\delta\xi(r)\}rr'drdr'}{\left[\int_{d/2}^{\infty}e^{ikr}\xi(r)dr\right]^2} -2\frac{\int_{d/2}^{\infty}\int_{d/2}^{\infty}G(0,r,r')\xi(r)\xi(r')rr'drdr'}{\left[\int_{d/2}^{\infty}e^{ikr}\xi(r)dr\right]^3} \times\left\{\int_{d/2}^{\infty}e^{ikr}\delta\xi(r)dr\right\} =\frac{2\left\{-(\zeta_0/8\pi)\int_0^h I(z)dz\right\}\int_{d/2}^{\infty}e^{ikr}\xi(r)dr}{\left[\int_{d/2}^{\infty}e^{ikr}\xi(r)dr\right]^2} -2\mu\frac{\int_{d/2}^{\infty}e^{ikr}\delta\xi(r)dr}{\left[\int_{d/2}^{\infty}e^{ikr}\xi(r)dr\right]}=0.$$

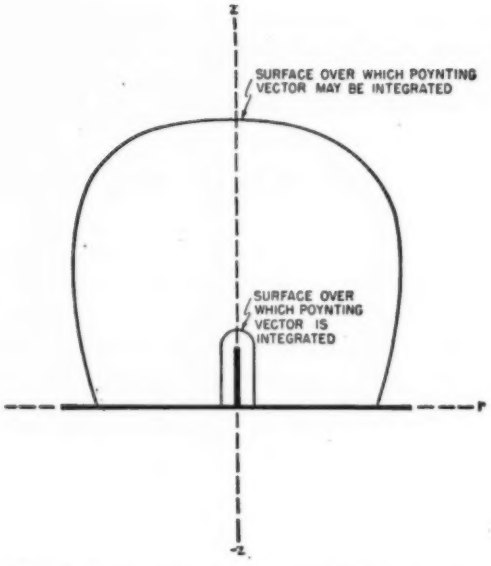


FIG. 3. Surfaces for Poynting vector integration.

Conversely, if μ in Eq. (11) is required to be stationary, it may be shown that the original integral equation (9) is obtained.

IV. EXPRESSION FOR THE INPUT IMPEDANCE

The purpose of this section is to obtain an expression for the input impedance, Z, of the antenna as a function of $\xi(r)$. Let σ denote a surface which surrounds the antenna, such as indicated in Fig. 3. Applying Poynting's theorem to this surface, one obtains the following results:

$$\mathcal{P}=\frac{1}{2}\int_\sigma (\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{n}d\sigma,$$

where real part $\mathcal{P}$ = mean power being radiated by the antenna, imaginary part $\mathcal{P}=2\omega\times$ difference between mean electric and magnetic energies inside σ, $\mathbf{n}$ is the unit outward normal vector to σ.

The expression developed for the fields, Eq. (7), is based on the assumption of an infinitesimally thin antenna and hence is only valid at large distances from the antenna. Therefore, if σ is taken to be sufficiently far away from the antenna, $\mathcal{P}$ becomes

$$\mathcal{P}=\frac{1}{2}\int_\sigma (\mathbf{E}^\infty+\mathbf{E})\times(\mathbf{H}^\infty+\mathbf{H})\cdot\mathbf{n}d\sigma.$$

Poynting's theorem is based on the maxwell equations. The fields, $\mathbf{E}^\infty+\mathbf{E}$, $\mathbf{H}^\infty+\mathbf{H}$, satisfy the maxwell equations and correctly represent the power being radiated when the surface σ is at a considerable distance from the antenna. Accordingly, these fields must also represent correctly the power being radiated if the surface σ is taken to be close to the antenna. In particular, if σ is allowed to shrink until it just surrounds the antenna, it is seen that

$$\mathcal{P}=\lim_{r\to 0}\frac{1}{2}\int_0^h\int_0^{2\pi}(E_z^\infty+\bar{E}_z)(H_\phi^\infty+\bar{H}_\phi)^*(\mathbf{z}\times\boldsymbol{\phi})\cdot\mathbf{r}rd\phi dz$$

$$=-\lim_{r\to 0}\frac{2\pi r}{2}\int_0^h(E_z^\infty+\bar{E}_z)(H_\phi^\infty+\bar{H}_\phi)^*dz. \qquad (12)$$

The results of this section may be obtained by taking σ to be a large spherical surface. This approach, however, involves more algebra than the one being used.

The quantity $\mathcal{P}$ is connected to the input impedance by

$$\mathcal{P}=\tfrac{1}{2}ZI(0)I^*(0).$$

$I(0)$ is the current at the driving point of the antenna. Inserting this expression for $\mathcal{P}$, Eq. (12) may be written in the form

$$ZI(0)I^*(0)=-\lim_{r\to 0}2\pi r\left\{\int_0^h(E_z^\infty+\bar{E}_z)(H_\phi^\infty+\bar{H}_\phi)dz\right\}.$$

Now, from Eq. (3), one obtains

$$\lim_{r\to 0}(H_\phi^\infty+\bar{H}_\phi)=\lim_{r\to 0}H_\phi^\infty=\lim_{r\to 0}\frac{I(z)}{2\pi r},\quad 0\leqq z\leqq h.$$

$\bar{H}_\phi$ does not contribute to this expression, since changes in radial currents on the ground screen (represented by $\bar{H}_\phi$) do not affect the distribution of current along the vertical antenna. This can readily be shown from the integral representation, 7, for $\bar{H}_\phi$.

Hence, it follows that

$$ZI(0)I^*(0)=-\int_0^h[E_z^\infty+\bar{E}_z]_{r=0}I^*(z)dz. \qquad (13)$$

Let Z_0 denote the impedance of the antenna over an infinite screen. Directly from the definition of $\bar{E}_z$, or from Eq. (7), it is apparent that $\bar{E}_z$ vanishes for an infinite screen. Hence, one obtains

$$Z_0I(0)I^*(0)=\lim_{d\to\infty}\left\{-\int_0^h[E_z^\infty+\bar{E}_z]_{r=0}I^*(z)dz\right\}$$

$$=-\int_0^h[E_z^\infty]_{r=0}I^*(z)dz.$$

Therefore, the first integral appearing in Eq. (13) is connected with impedance over an infinite screen. When this information is inserted, Eq. (13) becomes

$$(Z-Z_0)I(0)I^*(0)=-\int_0^h[\bar{E}_z]_{r=0}I^*(z)dz. \qquad (14)$$

Thus, it is seen that the effects of the infinite screen subtract out yielding directly an expression for the change of the impedance, $Z-Z_0$. Inserting the value for

$[\bar{E}_z]_{r=0}$ in terms of $[\bar{H}_\phi]_{r=0}$, one obtains

$$(Z-Z_0)I(0)I^*(0) = (\zeta_0/ik)\int_0^h [(1/r)(\partial/\partial r)r\bar{H}_\phi]_{r=0} I^*(z)dz. \quad (15)$$

When the value for $\bar{H}_\phi$ given by Eq. (7) (15) becomes

$$(Z-Z_0)I(0)I^*(0) = -2\int_0^h \int_{d/2}^\infty [(1/r)(\partial/\partial r)rG(0,r,r')]_{r=0} \times \xi(r')I^*(z)r'dr'dz. \quad (16)$$

Using the integral representation of $G(z, r, r')$ from Eq. (6) and noting that

$$(1/r)(\partial/\partial r)rJ_1(\lambda r) = -(1/\lambda)(1/r)(\partial/\partial r)r(\partial/\partial r)J_0(\lambda r) = \lambda J_0(\lambda r),$$

we see that

$$\begin{aligned}
&[(1/r)(\partial/\partial r)rG(z,r,r')]_{r=0} \\
&= \left[\frac{1}{r}\frac{\partial}{\partial r}\frac{i}{2}\int_0^\infty J_1(\lambda r)J_1(\lambda r') \times \exp(i[k^2-\lambda^2]^{\frac{1}{2}}|z|)\lambda d\lambda/[k^2-\lambda^2]^{\frac{1}{2}}\right]_{r=0} \\
&= -\tfrac{1}{2}i\left[\int_0^\infty J_0(\lambda r)J_1(\lambda r') \times \exp(i[k^2-\lambda^2]^{\frac{1}{2}}|z|)\lambda d\lambda/[k^2-\lambda^2]^{\frac{1}{2}}\right]_{r=0} \\
&= -\tfrac{1}{2}i\int_0^\infty J_1(\lambda r')\exp(i[k^2-\lambda^2]^{\frac{1}{2}}|z|)\lambda d\lambda/[k^2-\lambda^2]^{\frac{1}{2}} \\
&= -\tfrac{1}{2}i(\partial/\partial r')\int_0^\infty J_0(\lambda r') \times \exp(i[k^2-\lambda^2]^{\frac{1}{2}}|z|)\lambda d\lambda/[k^2-\lambda^2]^{\frac{1}{2}}.
\end{aligned}$$

It can be shown that[6]

$$\frac{\exp[ik(z^2+r^2)^{\frac{1}{2}}]}{(r^2+z^2)^{\frac{1}{2}}} = i\int_0^\infty J_0(\lambda r) \times \exp(i[k^2-\lambda^2]^{\frac{1}{2}}|z|)\lambda d\lambda/[k^2-\lambda^2]^{\frac{1}{2}},$$

and hence that

$$[(1/r)(\partial/\partial r)rG(z,r,r')]_{r=0} = -\tfrac{1}{2}(\partial/\partial r')\exp[ik(r'^2+z^2)^{\frac{1}{2}}]/(r'^2+z^2)^{\frac{1}{2}}.$$

Now, in Eq. (16), $kr' \geq kd/2 \gg 1$, and $r'^2 \gg k^2 \geq z^2$. The following approximation may accordingly be made:

$$[(1/r)(\partial/\partial r)rG(z,r,r')]_{r=0} \cong -\tfrac{1}{2}(\partial/\partial r')e^{ikr'}/r' \cong -ike^{ikr'}/2r'.$$

[6] Reference 4, pp. 575–576.

When this is inserted into Eq. (16), it becomes

$$(Z-Z_0)I(0)I^*(0) = ik\left\{\int_0^h I^*(z)dz\right\}\left\{\int_{d/2}^\infty e^{ikr'}\xi(r')dr'\right\}.$$

Eliminating the quantity

$$\left\{\int_{d/2}^\infty e^{ikr'}\xi(r')dr'\right\}$$

in favor of the parameter μ defined by Eq. (10), one obtains

$$(Z-Z_0)I(0)I^*(0) = (\zeta_0 k/8\pi i)(1/\mu)\left\{\int_0^h I(z)dz\right\}\left\{\int_0^h I^*(z)dz\right\};$$

or, solving for $Z-Z_0$, one obtains

$$Z-Z_0 = (\zeta_0/8\pi ik\mu)\left|k\int_0^h [I(z)/I(0)]dz\right|^2. \quad (17)$$

The utility of formulating the problem of the finite screen in this manner is already becoming apparent. If the value of μ is calculated from the variational expression (14), it depends only on the function $\xi(r)$. However, for a large screen it can be seen directly from Eq. (9), or from physical considerations, that the magnitude of $\xi(r)$ depends on the driving current but its shape does not. But the variational expression (14) for μ is independent of the magnitude of $\xi(r)$. Hence, the factor

$$\left|k\int_0^h [I(z)/I(0)]dz\right|^2$$

represents the entire dependence of $Z-Z_0$, for large screens, on the driving current. Evaluation of the quantity μ will then complete the picture by indicating the dependence of $Z-Z_0$ on the diameter of the screen.

V. EVALUATION OF μ

For convenience, the variational expression for μ [Eq. (11)] is now repeated.

$$\mu = \frac{\int_{d/2}^\infty \int_{d/2}^\infty G(0,r,r')\xi(r)\xi(r')rr'drdr'}{\left[\int_{d/2}^\infty e^{ikr}\xi(r)dr\right]^2}.$$

This expression for μ has the property of being stationary with respect to small changes of $\xi(r)$ about the correct value of $\xi(r)$. Hence, if an approximate value for $\xi(r)$ is used, a very good value for μ is obtained. Moreover, since the variational expression is independent of the over-all amplitude of $\xi(r)$, the function inserted need only resemble the correct $\xi(r)$ in its shape.

In order to see what function should be chosen, it is worth while to remember the Sommerfeld solution to the problem of the diffraction of a plane wave by an infinite half-screen. In this problem the electric field in the plane of the screen, which is analogous to $\xi(r)$, is given for small x by

$$E_x=(e^{ikx}/x^{\frac{1}{2}})\{A_0+A_1x+A_2x^2+\cdots\},\quad A_0\neq 0,$$

where $x=0$ represents the edge of the screen.

Since the problem of an antenna over a finite screen is in reality the problem of diffraction of waves around a large circular screen, it would be logical to choose for $\xi(r)$,

$$\xi(r)=e^{ikr}f(r)/(r-d/2)^{\frac{1}{2}},\tag{18}$$

where $f(r)$ is a slowly varying function of r, finite at $r=d/2$. Since $\xi(r)$ approaches e^{ikr}/r^2 as r approaches infinity, a logical choice for $\xi(r)$ would be $1/r^{\frac{1}{2}}$. However, it will be seen later that it is not even necessary to specify $f(r)$ in order to obtain a good value for μ.

This representation for $\xi(r)$ may also be obtained from the exact expression for $\xi(r)$ given by the series of spheroidal functions obtained by Leitner and Spence.[3] This series has exactly the character indicated by Eq. (18).

However, it is not at all necessary to refer to other work to obtain Eq. (18). The variational formulation will give this directly. Thus, if the very general function

$$\xi(r)=e^{ik'r}f(r)/(r-d/2)^{\nu}$$

is substituted into the variational expression, the resulting μ is a function of both ν and k'. Since this expression for μ is stationary, the best values of ν and k' are obtained by setting

$$\partial\mu/\partial k'=0,\quad \partial\mu/\partial\nu=0.$$

The procedure indicated may actually be performed although the results involve complicated integrals of bessel functions. The values of ν and k' are found to be $\nu=\frac{1}{2}$ and $k'=k$. Thus this procedure leads back to Eq. (18) and so it is certain that Eq. (18) will give the correct value of μ for large screens.

Inserting Eq. (18) into the variational expression for μ, one obtains

$$\mu=\frac{\displaystyle\int_{d/2}^{\infty}\int_{d/2}^{\infty}(rr')^{\frac{1}{2}}G(0,r,r')e^{ik(r+r')}\left[f(r)f(r')/(r-d/2)^{\frac{1}{2}}(r'-d/2)^{\frac{1}{2}}\right](rr')^{\frac{1}{2}}drdr'}{\left[\displaystyle\int_{d/2}^{\infty}e^{2ikr}f(r)dr/(r-d/2)\right]^2}.$$

Performing an integration by parts on the integral appearing in the denominator of the above expression,

$$\int_{d/2}^{\infty}e^{2ikr}\left[f(r)/(r-d/2)^{\frac{1}{2}}\right]dr=-f(r)\int_r^{\infty}\frac{e^{2ikt}}{(t-d/2)^{\frac{1}{2}}}dt\Bigg|_{d/2}^{\infty}+\int_{d/2}^{\infty}\left[\int_r^{\infty}\frac{e^{2ikt}}{(t-d/2)^{\frac{1}{2}}}dt\right]\frac{df(r)}{dr}dr$$

$$=f\left(\frac{d}{2}\right)\int_{d/2}^{\infty}\frac{e^{2ikt}}{(t-d/2)^{\frac{1}{2}}}dt+\int_{d/2}^{\infty}\left[\int_r^{\infty}\frac{e^{2ikt}}{(t-d/2)^{\frac{1}{2}}}dt\right]\frac{df(r)}{dr}dr.$$

The integral involving $df(r)/dr$ can be shown to be much less than the integrated term, since $f(r)\gg(1/k)df(r)/dr$. Thus, for example, if $f(r)$ were chosen to be $1/r^{\frac{1}{2}}$, then $df(r)/dr=-\frac{3}{2}(1/r^{5/2})$. Since $kr\geq k(d/2)\gg 1$, then $1/r^{\frac{1}{2}}\gg-\frac{3}{2}(1/kr^{5/2})$. Hence, the integral involving $df(r)/dr$ may be neglected for large screens. Using $r^{\frac{1}{2}}f(r)$ as the slowly varying function, the integral in the numerator may be similarly treated. This yields, for μ,

$$\mu=\frac{(d/2)f(d/2)f(d/2)\displaystyle\int_{d/2}^{\infty}\int_{d/2}^{\infty}(rr')^{\frac{1}{2}}G(0,r,r')\left[e^{ik(r+r')}/(r-d/2)^{\frac{1}{2}}(r'-d/2)^{\frac{1}{2}}\right]drdr'}{f^2(d/2)\left[\displaystyle\int_{d/2}^{\infty}\left[e^{2ikr}/(r-d/2)^{\frac{1}{2}}\right]dr\right]^2}$$

$$=\frac{d}{2}\frac{\displaystyle\int_{d/2}^{\infty}\int_{d/2}^{\infty}(rr')^{\frac{1}{2}}G(0,r,r')\left[e^{ik(r+r')}/(r-d/2)^{\frac{1}{2}}(r'-d/2)^{\frac{1}{2}}\right]drdr'}{\left[\displaystyle\int_{d/2}^{\infty}\left[e^{2ikt}/(t-d/2)^{\frac{1}{2}}\right]dt\right]^2}.$$

The integral in the denominator may readily be evaluated. Making the substitution $t'=t-d/2$, one obtains

$$\int_{d/2}^{\infty}[e^{2ikt}/(t-d/2)^{\frac{1}{2}}]dt=e^{ikd}\int_{0}^{\infty}e^{2ikt'}dt'/(t')^{\frac{1}{2}}$$
$$=(\pi/2k)^{\frac{1}{2}}e^{ikd+i(\pi/4)} \quad (19)$$

and hence

$$\mu=(kd/\pi i)e^{-2ikd}\int_{d/2}^{\infty}\int_{d/2}^{\infty}(rr')^{\frac{1}{2}}G(0,r,r')$$
$$\times[e^{ik(r+r')}/(r-d/2)^{\frac{1}{2}}(r'-d/2)^{\frac{1}{2}}]drdr'.$$

Since kr, $kr'\gg 1$, the function $G(0,r,r')$ may be simplified by replacing the bessel functions in its integral representation by their asymptotic values. The resulting integral can be evaluated. For details the reader is referred to the author's original paper.[7]

The value of μ so obtained is given by

$$\mu=\tfrac{1}{4}de^{-ikd}\{1+e^{ikd+i(3\pi/4)}/(2\pi kd)^{\frac{1}{2}}\}.$$

Returning to the expression for $Z-Z_0$ [Eq. (17)] and inserting this value for μ gives

$$Z-Z_0=\frac{\zeta_0}{2\pi kdi}e^{ikd}\left|k\int_0^h\frac{I(z)}{I(0)}dz\right|^2\left[1+\frac{e^{ikd+i(3\pi/4)}}{(2\pi kd)^{\frac{1}{2}}}\right]^{-1}.$$

The inequality $kd\gg 1$ is intended to mean that kd is of the order of 60 (or, equivalently, a screen of greater than 10 wavelengths in diameter). Hence, it follows that $1/(2\pi kd)^{\frac{1}{2}}\leq 0.05$. Therefore, the last factor in the expression for $Z-Z_0$ may be dropped, as it is quite negligible. This gives as a final result

$$Z-Z_0=(\zeta_0/2\pi kdi)e^{ikd}\left|k\int_0^h[I(z)/I(0)]dz\right|^2. \quad (20)$$

VI. INTERPRETATION OF RESULTS; COMPARISON WITH EXPERIMENT

Inserting $k=2\pi/\lambda$, where λ is the free-space wavelength, and the numerical value $\zeta_0=120\pi$ ohms, the expression for $Z-Z_0$ [Eq. (20)] becomes

$$Z-Z_0=\frac{9.55}{i(d/\lambda)}e^{2\pi i(d/\lambda)}\left|k\int_0^h\frac{I(z)}{I(0)}dz\right|^2 \text{ ohms.} \quad (21)$$

Separating into real and imaginary parts gives

$$R-R_0=\frac{9.55}{(d/\lambda)}\sin 2\pi\frac{d}{\lambda}\left|k\int_0^h\frac{I(z)}{I(0)}dz\right|^2,$$

$$X-X_0=\frac{9.55}{(d/\lambda)}\cos 2\pi\frac{d}{\lambda}\left|k\int_0^h\frac{I(z)}{I(0)}dz\right|^2.$$

Thus, both the resistance and reactance are damped periodic functions of d with period λ. Plotted on the complex plane, $Z-Z_0$ is a spiral of slowly decreasing radius. Since the factor $|k\int_0^h[I(z)/I(0)]dz|^2$ is dimensionless, Eq. (21) may be divided through by it and a "universal" curve obtained. This appears in Fig. 4.

The dimensionless factor $|k\int_0^h[I(z)/I(0)]dz|^2$ represents the dependence of $Z-Z_0$ on the current. For $kh\leq\frac{3}{4}\pi$, it is a well-known fact that the current on the antenna, for any diameter of antenna, can be approximated by

$$I(z)=I(0)[\sin k(h-z)/\sin kh], \quad kh\leq\tfrac{3}{4}\pi. \quad (22)$$

Using (22), one obtains

$$\left|k\int_0^h[I(z)/I(0)]dz\right|^2=[(1-\cos kh)/\sin kh]^2,$$
$$kh\leq\tfrac{3}{4}\pi. \quad (23)$$

It is seen that the factor $|k\int_0^h[I(z)/I(0)]dz|^2$ equals 1 for the dipole antenna, $kh=\pi/2$. Hence, the "universal" curve of Fig. 4 may also be given the interpretation of being the change of impedance for a dipole antenna.

Examining Eq. (23), we see that this factor is small for small kh and becomes progressively larger as kh is increased. Near antiresonance, $kh\cong\pi$, the current on the antenna is quite dependent on the diameter of the antenna as well as its length. Hence, to evaluate this factor near antiresonance would require the substitution of a more accurate current distribution than Eq. (22)—such as those given by King,[8,9] for example. Since this factor is very large near antiresonance, the impedance change, $Z-Z_0$, may easily get as large as 40 ohms. However, the impedance itself is large near antiresonance and this is only a small percentage variation.

To the best of the author's knowledge, the only experimental measurements with which it is possible to compare these results is the aforementioned work of Meier and Summers,[1] who verified the fact that $Z-Z_0$, plotted on the complex plane, is a spiral of slowly decreasing radius. Moreover, the fact that this spiral has a "period," or completes one revolution, as d is increased by a wavelength, is in substantial agreement with experimental results.

Qualitatively, the factor $|k\int_0^h[I(z)/I(0)]dz|^2$ appears to represent very satisfactorily the experimental data on antenna length. The measurements of Meier and Summers unfortunately do not extend to sufficiently large values of d to make possible a quantitative check of this dependence.

The only data of Meier and Summers with which it appears possible to make a quantitative comparison are

[7] J. E. Storer, "The impedance of an antenna over a large circular screen," Technical Report No. 119, Cruft Laboratory, Harvard University, November 30, 1950.

[8] R. King, "Graphical representation of the characteristics of cylindrical antennas," Technical Report No. 20, Cruft Laboratory, Harvard University, October 1, 1947.

[9] D. Middleton and R. King, J. Appl. Phys. **17**, 273–284 (1946).

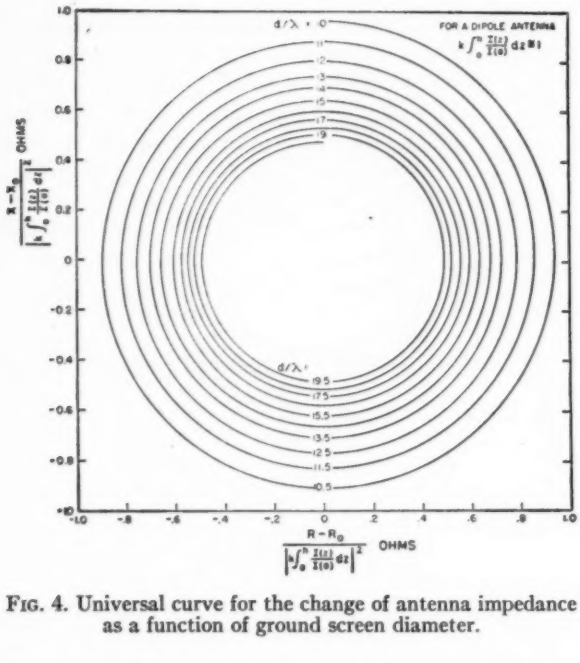


FIG. 4. Universal curve for the change of antenna impedance as a function of ground screen diameter.

for an antenna length of $h/\lambda=0.224$ and $4<d/\lambda<6$. The inequality $d^2\gg h^2$ is satisfied. However, the inequality $kd\gg 1$ is meant to imply that $kd=2\pi d/\lambda$ is of the order of 60. But for $d/\lambda\cong 5$, $kd\cong 30$, and the inequality is not satisfied. However, the agreement should be close.

Since $kh\leq\frac{3}{4}\pi$, Eq. (23) may be used; and the impedance change for this length of antenna becomes

$$Z-Z_0=\frac{9.55}{i(d/\lambda)}e^{2\pi i(d/\lambda)}\left[\frac{1-\cos 2\pi(0.224)}{\sin 2\pi(0.224)}\right]^2$$
$$=[6.90/(d/\lambda)]e^{2\pi i(d/\lambda)-i(\pi/2)}. \qquad (24)$$

The phase of $Z-Z_0$ given by Eq. (24) agrees to within experimental error. Taking magnitudes on both sides of Eq. (24), one obtains

$$(d/\lambda)|Z-Z_0|=6.90.$$

Experimentally, the following is measured:

d/λ	$(d/\lambda)\lvert Z-Z_0\rvert$
4.5	9.1 ± 0.5
5.0	8.4 ± 0.5
5.5	7.8 ± 0.5
Theoretical	6.90.

The agreement is already close. It is not difficult to believe that when the condition $kd\gg 1$ is satisfied, say, for $d/\lambda=10$, the theoretical and measured values will agree.

VII. CONCLUSION

This paper has derived a simple formula for the change of the input impedance of an antenna as a function of the diameter of the ground plane. As will be shown in a later paper, it is also possible to obtain from this formulation an expression for the fields over a finite ground plane.

However, these results are not of prime interest. Usually all that is desired is to know whether the screen is large enough so that for practical purposes it may be considered an infinite screen. The formula for the impedance change [Eq. (21)] may be used to determine this. To illustrate, consider the typical example of a quarter-wavelength antenna over a 10-wavelength ground plane. From Eq. (21) it is found that the impedance change for this case is about 1 ohm. Since the impedance of this antenna is of the order of 40 ohms, the use of a 10-wavelength ground plane instead of an infinite one produces a change of $2\frac{1}{2}$ percent. Whether or not this change may be regarded as a negligible effect depends upon the type of measurements being made.

ACKNOWLEDGMENT

The author wishes to thank Dr. R. King under whose guidance this paper was prepared. He also wishes to acknowledge the financial assistance given him in the form of a fellowship by the Atomic Energy Commission.

Expected Number of Crossings of Axis by Linearly Increasing Function Plus Noise*

FRANK W. LEHAN
Chief of Telecommunication, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received March 7, 1951)

The expected number of times the function $e(t)=at+e_N(t)$ crosses the t axis in a positive direction prior to any given time is calculated, where $e_N(t)$ is a random noise function of arbitrary power spectrum. The solution is given as the product of two functions. Curves are presented for the two functions.

I. INTRODUCTION AND SUMMARY

MANY situations arise in which a signal voltage, which is approximately a linearly increasing function of time, is caused to trigger some device when the voltage reaches a preset value. For example, the functioning of a relay or thyratron in an on-off control mechanism frequently may be resolved into an occurrence of this kind. Upon the signal voltage there is usually superimposed an unavoidable fluctuating noise voltage which causes an error in the time at which triggering occurs.

It is of interest to find the probability distribution of the error. Although this problem has not been solved, the results of a related but simpler problem are given in the present discussion. The problem to be solved is to find the expected number of crossings of the axis (in a positive direction), prior to a given time T, by a voltage which is the sum of a linearly increasing function of time $e_s(t)=at$ and a random noise voltage of arbitrary power spectrum $e_N(t)$. This solution is a generalization of the problem of finding the expected number of times per second a noise voltage crosses a fixed arbitrary reference level.[1] The situation is illustrated graphically in Fig. 1. The solution to this problem is presented in Sec. II, and details of the solution are given in the Appendix.

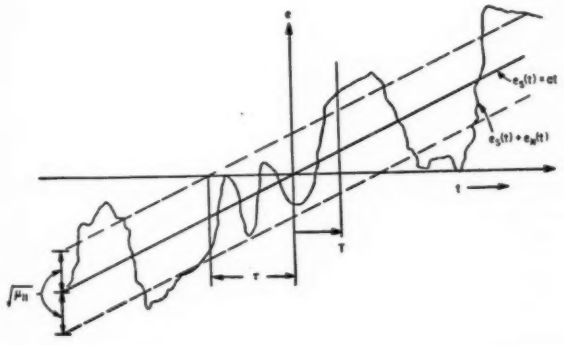


FIG. 1. Signal-plus-noise voltage in region of crossing of axis.

II. SOLUTION TO PROBLEM

The expected number of crossings (from the negative to the positive side of the time axis) prior to time T is found to be given by

$$C=\Gamma(\gamma)\Phi(\phi), \tag{1}$$

where

$$\Phi(\phi)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\phi}e^{-x^2/2}dx \tag{2}$$

is the normal probability distribution function.[2] It is plotted in Fig. 2. As will be seen, $\Phi(\phi)$ gives the fraction of the total expected number of crossings that may be expected to be completed up to a given time.

The function $\Gamma(\gamma)$ is given by

$$\Gamma(\gamma)=[\Phi'(\gamma)/\gamma]+\Phi(\gamma), \tag{3}$$

where

$$\Phi'(\gamma)=d\Phi(\gamma)/d\gamma=(2\pi)^{-\frac{1}{2}}e^{-\gamma^2/2} \tag{4}$$

is the normal probability density function. The function $\Gamma(\gamma)$ is plotted in Fig. 3. It is the total expected number of crossings.

The variable ϕ is a normalization of the variable T and is given by

$$\phi=T/\tau, \tag{5}$$

where

$$\tau=\mu_{11}^{\frac{1}{2}}/a, \tag{6}$$

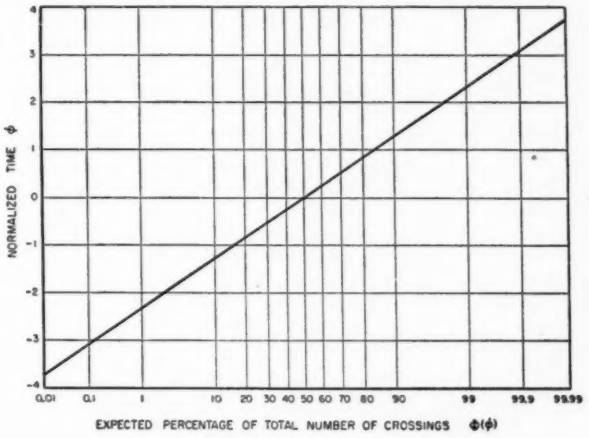


FIG. 2. Plot of $\Phi(\phi)$ *vs* ϕ.

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. DA-04-495-ORD 18, sponsored by the U. S. Army Ordnance Department.

[1] S. O. Rice, Bell System Tech. J. **23**, 282–332 (1944); **24**, 46–156 (1945).

[2] Harald Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946), p. 208.

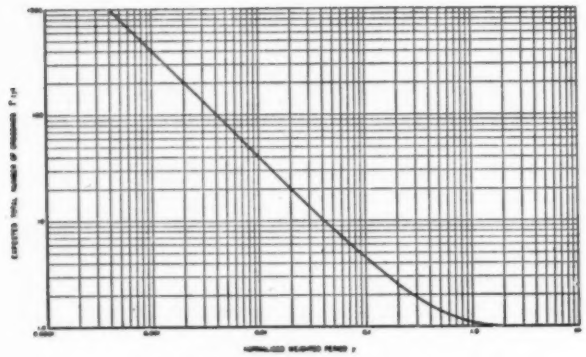


FIG. 3. Plot of $\Gamma(\gamma)$ *vs* γ.

and where a is the slope of the $e_s(t)$ line and $\mu_{11}^{\frac{1}{2}}$ is the rms level of the noise. The parameter τ is indicated in Fig. 1.

The variable γ is a normalization of the weighted average period of the noise and is given by

$$\gamma = T_c/\tau, \tag{7}$$

where

$$T_c^2 = \int_0^\infty dW(f)/4\pi^2 \int_0^\infty f^2 dW(f) \tag{8}$$

and where, assuming a 1-ohm resistance, $W(f)$ is the total noise power between zero frequency and frequency f. (This power would be read by a thermal power meter at the output of a low pass filter of cut-off frequency f, if the noise were supplied to its input.) The integral in Eq. (8) is to be taken in the Stieltjes sense.

APPENDIX. METHOD OF SOLUTION

As a first step in finding the expected number of crossings prior to time T, it is desired to determine the probability that the noise voltage $e_N(t)$ will pass, in the time interval (t_1, t_1+dt),† through a straight line whose slope is $-a$ and whose value (at $t=t_1$) is e_1. This time interval is taken short enough to assure that the segment of $e_N(t)$ may be considered a straight line. An examination of Fig. 4 shows that this condition places certain restrictions on the variables ξ and η, where ξ is the value of e_N at $t=t_1$, and η is the slope of e_N at $t=t_1$. As was shown by Rice,[1] both ξ and η are random variables. The variable (ξ, η) is a joint random variable

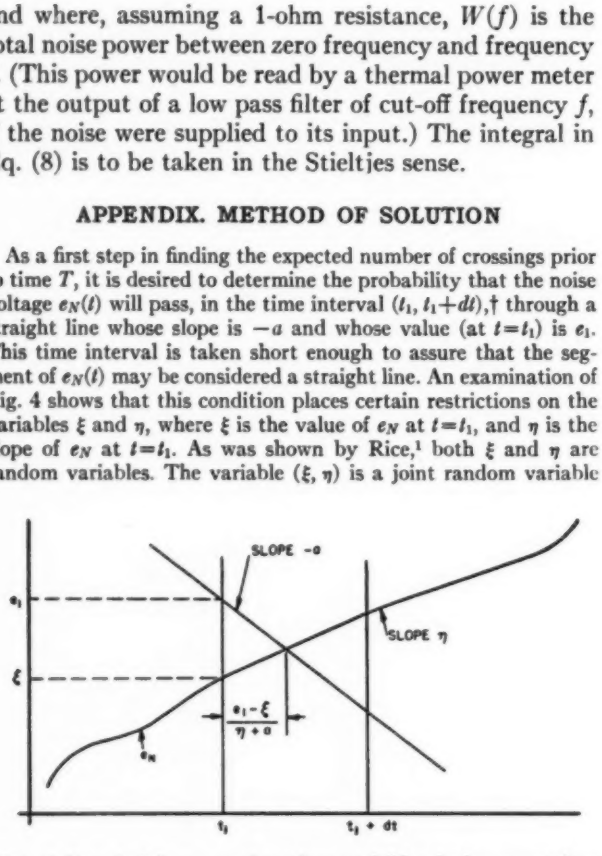


FIG. 4. Permissible range of random variables during a crossing.

† The noise voltage $e_N(t)$ passing through the line of the slope $-a$ is equivalent to the signal plus the noise voltage passing through the value zero.

with a certain continuous probability density function $f(\xi, \eta)$. Thus, (ξ, η) is seen to be restricted to a certain region S_1 of the (ξ, η) plane. The probability of this occurrence is given by

$$P(S_1) = \iint f(\xi,\eta)\, d\xi d\eta. \tag{A1}$$

From Fig. 4, it is obvious that S_1 is described by the relation,

$$t_1 \leqq t_1 + (e_1 - \xi)/(\eta + a) \leqq t_1 + dt \tag{A2}$$

or

$$(e_1 - a\,dt) - \eta\,dt \leqq \xi \leqq e_1. \tag{A3}$$

The region S_1 is pictured in Fig. 5.

Equation (A1) then becomes

$$P(S_1) = \int_{\eta=-a}^{\eta=\infty} d\eta \int_{\xi=e_1-(a+\eta)dt}^{\xi=e_1} f(\xi,\eta)d\xi. \tag{A4}$$

If dt is sufficiently small to make $(a+\eta)dt$ negligible except for all values of η such that the integral

$$\int_\eta^\infty d\eta \int_{\xi=e_1-(a+\eta)dt}^{\xi=e_1} f(\xi,\eta)d\xi$$

does not differ appreciably from zero,‡ then one obtains

$$P(S_1) = dt \int_{-a}^{\infty} (a+\eta)\, f(e_1,\eta)\, d\eta. \tag{A5}$$

Next, an expression for $f(\xi, \eta)$ is needed. It has been derived by

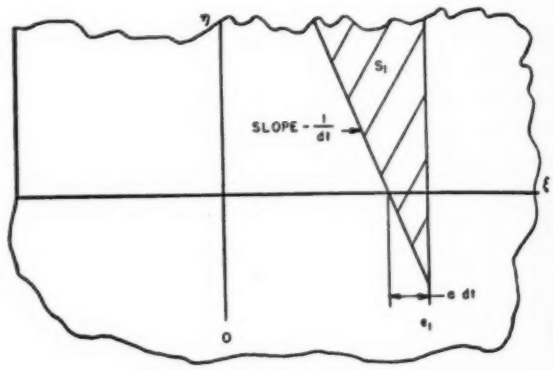


FIG. 5. Sketch of region S_1.

Rice[1] and is given in the following equation:

$$f(\xi,\eta) = \frac{1}{2\pi(\mu_{11}\mu_{22})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\frac{\xi^2}{\mu_{11}} + \frac{\eta^2}{\mu_{22}}\right)\right], \tag{A6}$$

where μ_{11} is the expected value[3] of ξ^2 and is given by

$$\mu_{11} = E(\xi^2) = \int_0^\infty dW(f), \tag{A7}$$

while

$$\mu_{22} = E(\eta^2) = 4\pi^2 \int_0^\infty f^2 dW(f). \tag{A8}$$

Note that ξ and η are independent, normally distributed random variables.

Equation (A5) now becomes

$$P(S_1) = dt \int_{a=\infty}^{\infty} (a+\eta) \frac{1}{2\pi(\mu_{11}\mu_{22})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\frac{e_1^2}{\mu_{11}} + \frac{\eta^2}{\mu_{22}}\right)\right] d\eta. \tag{A9}$$

Letting $e_1 = at$ and evaluating the integral in Eq. (A9), one obtains

$$P(S_1) = \frac{dt}{(2\pi\mu_{11})^{\frac{1}{2}}}\left[\left(\frac{\mu_{22}}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{a^2}{2\mu_{22}}\right) + a\Phi\left(\frac{a}{\mu_{22}^{\frac{1}{2}}}\right)\right] \times \exp\left(-\frac{a^2t^2}{2\mu_{11}}\right). \tag{A10}$$

‡ This is not a severe restriction for types of $f(\xi, \eta)$ likely to be encountered in noise phenomena.

[3] Reference 2, pp. 170–174.

The expected number C of crossings of the axis prior to time T is

$$C=\left[\left(\frac{\mu_{22}}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{a^2}{2\mu_{22}}\right)+a\Phi\left(\frac{a}{\mu_{22}^{\frac{1}{2}}}\right)\right]\frac{1}{(2\pi\mu_{11})^{\frac{1}{2}}}\times\int_{-\infty}^{t-T}\exp\left(-\frac{a^2t^2}{2\mu_{11}}\right)dt \quad \text{(A11)}$$

or

$$C=\left[\frac{1}{a}\left(\frac{\mu_{22}}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{a^2}{2\mu_{22}}\right)+\Phi\left(\frac{a}{\mu_{22}^{\frac{1}{2}}}\right)\right]\Phi\left(\frac{aT}{\mu_{11}^{\frac{1}{2}}}\right). \quad \text{(A12)}$$

Now let

$$(\mu_{11})^{\frac{1}{2}}/a=\tau, \quad \text{(A13)}$$

$$\mu_{22}=\mu_{11}/T_e^2, \quad \text{(A14)}$$

$$T/\tau=\phi, \quad \text{(A15)}$$

$$T_e/\tau=\gamma. \quad \text{(A16)}$$

Then one obtains

$$C=[(2\pi)^{-\frac{1}{2}}\gamma^{-1}e^{-\gamma^2/2}+\Phi(\gamma)]\Phi(\phi); \quad \text{(A17)}$$

or if

$$\Gamma(\gamma)=[\Phi'(\gamma)/\gamma]+\Phi(\gamma), \quad \text{(A18)}$$

then the final expression is obtained as follows:

$$C=\Gamma(\gamma)\Phi(\phi). \quad \text{(A19)}$$

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

Measurement of Velocity and Pressure of Gases in Rocket Flames by Spectroscopic Methods

F. P. Bundy, H. M. Strong, and A. B. Gregg
Research Laboratory, General Electric Company, Schenectady, New York

(Received April 12, 1951)

A method for measuring the gas flow velocity in hot high velocity exhaust flames is described. This method is based on the doppler shift of wavelength of the sodium or lithium spectral radiation from the flame when viewed at different angles relative to the flame axis. The same apparatus may be used to measure the gas pressure from the pressure shift of wavelength of the sodium radiation. A Fabry-Perot interferometer and camera assembly was used to determine the small shift in wavelengths. It was concluded that the absolute accuracy of any one velocity measurement by this method is about $\pm 1.5\times 10^4$ cm/sec, while that of any one pressure measurement is about ± 0.3 atmosphere. The method is therefore applicable to flames of very high velocity and to pressures of at least a few atmospheres.

I. INTRODUCTION

THE usual way of determining the gas velocities in the exhaust flames from rocket motors on test stands is to measure the mass flow rate of the reactants and the thrust force, and from these deduce the gas velocity. In a liquid-fuel liquid-oxidizer combination of reactants this requires the measurement of three quantities, viz., the fuel rate, the oxidizer rate, and the thrust force. As each measurement has its characteristic probable error there is a possibility of the deduced velocity having a considerable error. In the case of solid fuel motors an accurate measurement of the reactant flow rate is even more difficult.

The present paper describes a method developed by the authors for measuring the exhaust gas velocity directly from the flame by observing the doppler shift of wavelength of the Na or Li line spectrum radiation emitted from the flame at different angles. This method involves no physical disturbance of the flame and is entirely independent of the mass flow rate and thrust force, and thus gives a completely separate measurement of the gas velocity which can be compared to that obtained from the conventional method. In practice, the doppler shift method requires the use of very accurate optical equipment as well as considerable skill to operate it, and for this reason it is not proposed as a substitute for the routine method but rather as a special method to give an independent comparison value.

The same spectroscopic apparatus was also capable of measuring the small shifts of wavelength of Na D-lines associated with changes in pressure in the flame gases. In these measurements it was necessary to view the flame perpendicularly to eliminate any doppler effect. Calibration of the pressure shift in flame gases was carried out in the laboratory by observing the shift in wavelength of the same spectral lines from flames of known pressures. This method was used to measure the gas pressures at several places of interest in the flame.

In addition to the velocity and pressure results, some spectroscopic phenomena of interest to physicists and astrophysicists were observed. One of the most unusual effects was the wavelength shift of the absorption relative to the emission in the sodium lines owing to the differences in velocity and temperature of the different radial zones of the flame in the line of sight.

II. RESONANCE RADIATION FROM FLAMES

As the shifts in wavelength dealt with in this study are in general smaller than the widths of the spectral lines themselves, and as the widths of resonance-type spectral lines are influenced by conditions in the flame, the subject of line width will be analyzed briefly to give a background for experimental methods of keeping the line width as small as possible.

Any spectral line consists of a narrow band of wave-

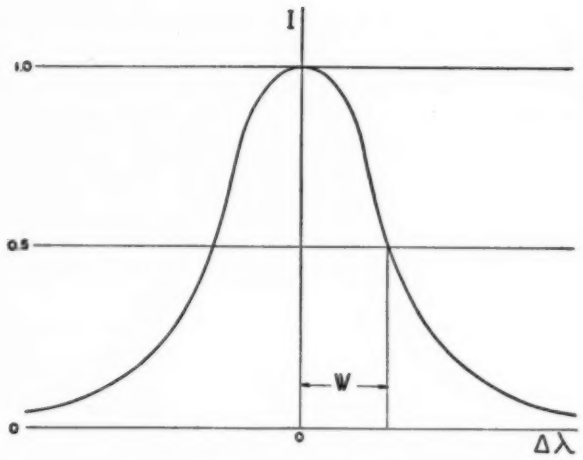


FIG. 1. Intensity distribution in a spectral line.

lengths in which the intensity drops off rapidly on both sides of the line center, as illustrated in Fig. 1. The wavelength interval between the line center and a half-intensity point is designated as W. The exact intensity contour of a resonance-type line depends upon the temperature and pressure of the flame gases, the physical thickness of the flame, and the concentration of the emitting element.

In a very thin flame where self-absorption effects are negligible the line width is controlled by the doppler effect from the thermal motion of the emitting atoms, and by the disturbing effect of collisions of the emitting atoms with neighboring molecules and atoms. The magnitude of the "doppler width" is given by the expression,

$$W_{\text{dopp}} = 1.18(\lambda/c)(RT/m)^{\frac{1}{2}}, \tag{1}$$

where λ is the wavelength, c the velocity of light, R the universal gas constant, T the absolute (Kelvin) temperature, and m the atomic weight of the emitting atoms. The intensity contour stemming from this effect is "gaussian" in form, and when normalized to have unity value at the line center may be expressed as the following "*shape factor*":

$$sf_{\text{dopp}} = \exp[-(\ln 2/W^2_{\text{dopp}})\langle\Delta\lambda\rangle^2], \tag{2}$$

where λ is the wavelength interval measured from the center of the line.

The magnitude of the "collision width" is given by the expression,

$$W_{\text{coll}} = (2\lambda^2/c)Nr^2p(\pi m'RT)^{-\frac{1}{2}}, \tag{3}$$

where N is the number of atoms per mole, r is the effective collision diameter of the emitting atom considered as a target for the flame gas molecules, p is the absolute pressure, and m' is the reduced atomic mass of an emitting atom and a flame gas molecule considered as a planetary pair. The *shape factor* associated with this collision effect is of the damped resonance type:

$$sf_{\text{coll}} = (W^2_{\text{coll}}/W^2_{\text{coll}} + \langle\Delta\lambda\rangle^2). \tag{4}$$

The extremities of this resonance-type contour are much higher than those of the gaussian-type, and for this reason the shape of the line changes with the temperature and pressure of the flame. In Fig. 2 the half-width of a sodium line caused by doppler and collision effects at one atmosphere absolute pressure is plotted against the absolute temperature. Also plotted is the half-width caused by these effects combined. It is interesting to note that above about 1600°K the half-width remains nearly constant.

The line shapes discussed above apply only to very thin flames in which self-absorption effects are very small. For flames of greater optical depth the self-absorption effect causes further broadening and change of shape of the spectral line. This results from the interplay of emission and absorption of the resonance radiation by the optically active atoms. This type of line broadening is quite sensitive to the exact shape, $sf\,(\Delta\lambda)$, of the basic line contours discussed above. It can be shown that the final intensity distribution in a resonance-type spectral line is given by the expression,

$$I(\Delta\lambda) = E_b[1-\exp(-zsf(\Delta\lambda)], \tag{5}$$

where E_b is the emissive power of a blackbody at the flame temperature, and z is the "optical depth" of the flame at the wavelength of the center of the line ($\Delta\lambda=0$). The unit of "optical depth" as used here is defined as that depth of flame gas required to reduce by primary absorption the intensity of an incident beam of monochromatic light having the wavelength of the center of the resonance line to exp (-1) of its initial value. The optical depth is proportional to the number of optically active atoms in the line of sight.

It has been determined experimentally that the optical depth per centimeter depth of flame is about $(1.5\times10^{-13})n$, where n is the number of sodium atoms

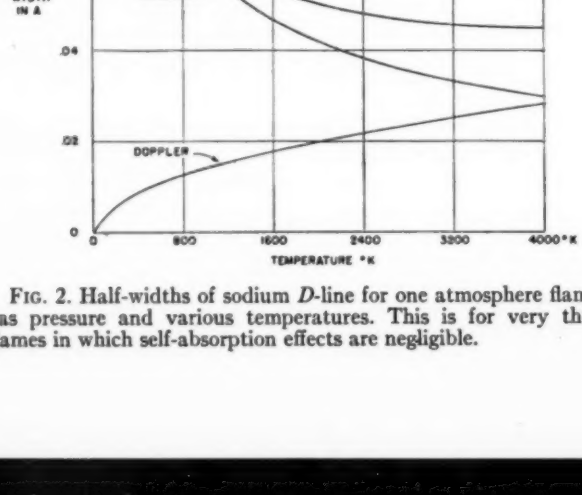


FIG. 2. Half-widths of sodium D-line for one atmosphere flame gas pressure and various temperatures. This is for very thin flames in which self-absorption effects are negligible.

per cm³. Thus

$$z=1.5\times10^{-13}nL, \qquad (6)$$

where L is the length of the optical path in the flame.

When the collision-type shape factor predominates (high pressure, low temperature) the half-width of the intensity distribution given in Eq. (5) increases much more rapidly with the optical depth than it does when the thermal doppler-type shape factor predominates (low pressure, high temperature). This effect is illustrated in Fig. 3 in which the half-width of a sodium D line is plotted against optical depth for flames at one atmosphere pressure and various temperatures. On this logarithmic plot the curves have a slope of very nearly one-half, which means that the line width increases about as the square root of the optical depth. It is obvious from this graph that the sodium vapor concentration must be kept quite low if the sodium D lines are to be sharp and narrow.

Another flame characteristic which can affect the shape and sharpness of the spectral lines of the resonance

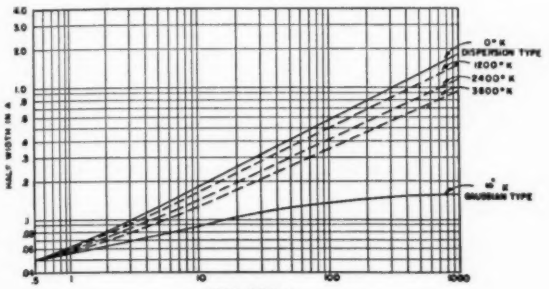


FIG. 3. Half-width of sodium D-line for one atmosphere flame gas pressure and various temperatures plotted against optical depth of flame.

type is the presence of zones of different temperatures. Consider, for example, a flame comprised of a hot core and relatively cool sheath as illustrated in Fig. 4. The intensity contours of a sodium D line emitted from such a flame for several different values of optical depth z are shown in Fig. 5. For measurement of small wavelength shifts the narrow line corresponding to $z=\frac{1}{2}$ is far more satisfactory than the broader dimpled ones corresponding to greater optical depths. A second point of practical interest which shows up in Fig. 5 is that in spite of the 8 to 1 variation in sodium density the maximum intensity in any of the line contours is about the same. This means that the D-lines from a $z=\frac{1}{2}$ flame can be photographed as readily as those from a $z=8$ flame in an optical instrument with enough resolving power and dispersion to display the line contour. In a light path 10 cm long an optical depth of $z=\frac{1}{2}$ corresponds to a sodium concentration of 3.3×10^{11} atoms per cm³. This is an exceedingly low concentration, barely enough to give the flame a faint yellow color. In alcohol-oxygen burning rockets there is usually more than enough contamination sodium in the alcohol

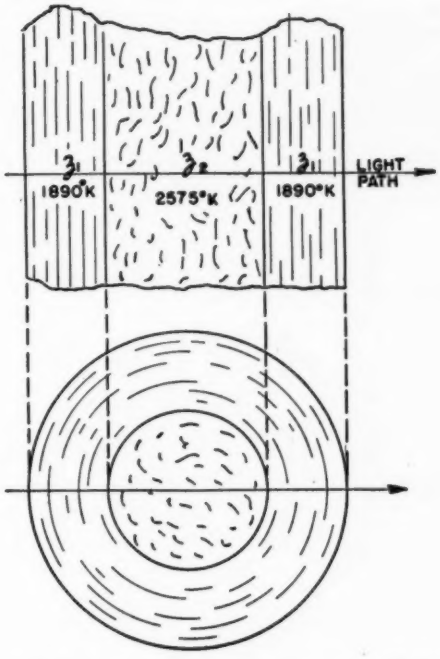


FIG. 4. Flame with hot core and relatively cool sheath.

fuel to give an optical depth of $\frac{1}{2}$ in a flame 10 cm in diameter.

III. INSTRUMENTATION

The shifts in wavelength to be measured were all less than 0.1 Angstrom unit, and to measure them accurately a spectroscopic instrument of high dispersion

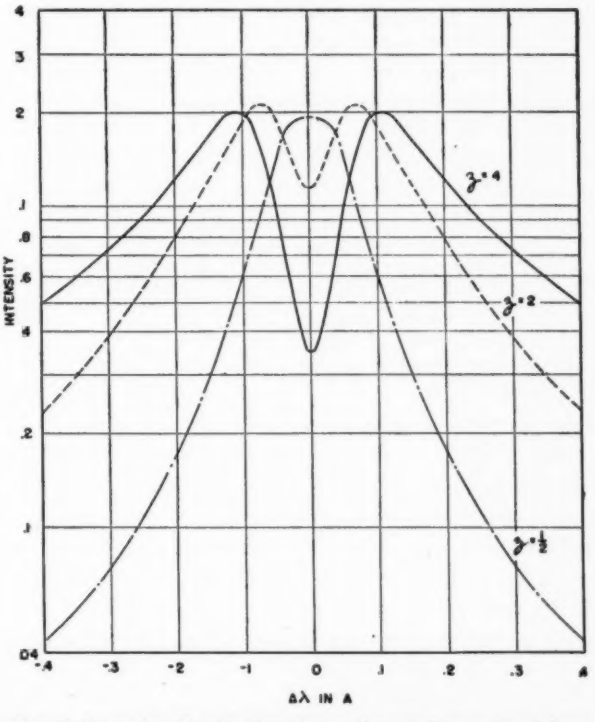


FIG. 5. Intensity distribution in a sodium D-line emitted from the flame illustrated in Fig. 4.

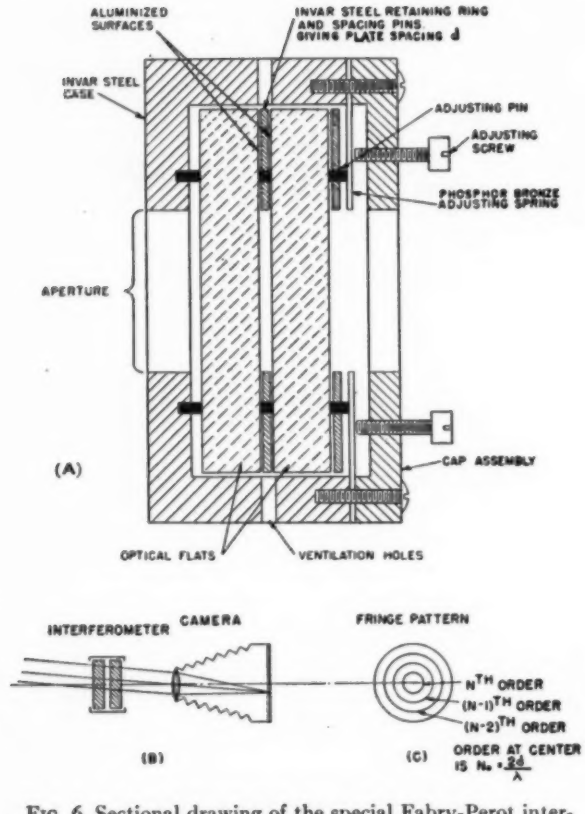


FIG. 6. Sectional drawing of the special Fabry-Perot interferometer, and the interferometer camera system.

and great resolving power was required. In addition, the instrument had to be small, compact, and capable of operating near a rocket test stand where the vibration conditions were quite severe.

A sectional drawing of the interferometer is shown in Fig. 6(A). The essential parts were two very accurate optical flats* which were held apart by three invar steel spacing pins. These pins held the plates parallel to an accuracy of 0.01 wavelength of sodium light so that the interference conditions were the same regardless of the distribution of intensity of illumination over the aperture area. The two inner faces of the optical flats were aluminized to the extent that each surface transmitted about 10 percent of the light incident upon it. Such an arrangement of parallel, partially transmitting mirrors has the property of transmitting monochromatic light only at certain angles to its optic axis, and hence when a camera set to focus incident parallel light is placed on the exit side of the interferometer as shown in Fig. 6(B) a pattern of concentric circular fringes is formed on the camera plate (Fig. 6(C)), each circular fringe representing a given order of interference. The diameters D of these fringes depend upon the wavelength of the light, the spacing d of the interferometer plates, and the focal length F of the camera in the following way:

$$D_n^2 = 4F^2[(2d/n\lambda)^2 - 1], \tag{7}$$

where n is the order of interference of the fringe. For the fringe of smallest diameter, n_1 is the next integer less than $2d/\lambda$; for the second fringe, n_2 is the second smaller integer, etc. If we let

$$x = (2d/\lambda) - n, \tag{8}$$

which is the absolute value of the difference of order of interference between a given fringe and the center of the pattern, it can be shown that to an accuracy better than 0.3 percent at 5° from the optic axis:

$$D_x^2 = (4F^2\lambda/d)x. \tag{9}$$

Thus, the difference of the squares of the diameters of two fringes $x=p$ and $x=q$ on a given interference pattern is simply

$$D_p^2 - D_q^2 = (4F^2\lambda/d)(p-q). \tag{10}$$

This relationship makes it possible to determine the constant $(4F^2\lambda)/d$ by merely measuring the diameters of two fringes on a given pattern.

Except for a slight spreading effect caused by the lack of resolving power, the distribution of intensity across a given fringe in the radial direction is that of the so-called monochromatic spectral line of which the light is composed. Thus each fringe is a repetition of the spectral line. In case the spectral line width exceeds the wavelength interval corresponding to the spacing from one fringe to the next, the extremities of adjacent fringes overlap to give a continuous background of intensity. This effect decreases the sharpness and contrast of the pattern and is another reason for attempting to keep the line width of the radiation from the flame as small as possible.

As the sodium D line radiation consists of two spectral lines, 5890A and 5896A, there are, in general, two sets of fringes in the interference pattern. For plate spacings which are multiples of 0.02906 cm, the two sets of fringes coincide, thereby increasing the intensity

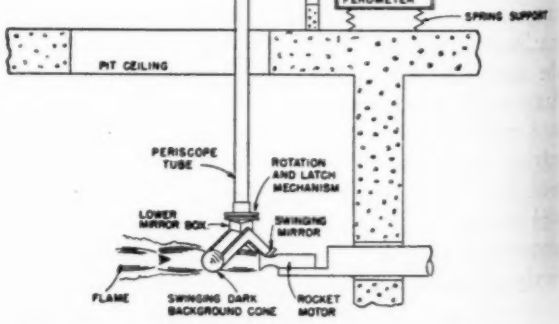


FIG. 7. Diagram of the optical setup at the rocket test stand.

* Made by the Perkin-Elmer Corporation, Glenbrook, Connecticut, flat over the aperture area to an accuracy of 1/50 wavelength of sodium light, and the two adjacent reflecting surfaces in the interferometer parallel within 1/100 wavelength.

of the fringes and improving the cleanness of the area between fringes.

For constant plate spacing an increase of wavelength of the light $\delta\lambda$ causes each fringe in the pattern to shrink toward the center in such a way that the order of interference of each decreases by an amount δx, which is equal to the change δn_0 of the order of interference $n_0 = 2d/\lambda$ at the center. Thus

$$\delta x = \delta n_0 = -(2d/\lambda^2)\delta\lambda. \quad (11)$$

The relation between the change of diameter of the fringes δD and the change of wavelength is obtained by differentiating Eq. (9),

$$2D\delta D = (4F^2\lambda/d)\delta x,$$

and substituting for δx the value in Eq. (11). One gets,

$$(\delta\lambda/\lambda) = -(D\delta D/4F^2). \quad (12)$$

By use of Eq. (10), this can be put in the form,

$$\delta\lambda/\lambda = -\{(\lambda/d)[(p-q)/(D_p^2 - D_q^2)]\}D\delta D, \quad (13)$$

from which the fractional wavelength change can be determined from measurements of the diameters of the fringes on the two patterns being compared. It is important to note that since the product $D\delta D$ is the same for all fringes in the pattern the small diameter fringes near the center give the greatest change in diameter and thus a more accurate indication of the shift. In the measurements discussed in this paper the shifts were always less than one fringe width so it was necessary to photograph two separate interference patterns corresponding to the two conditions giving the shift; then the fringe diameters of each pattern were measured, and from their differences and mean values the shift of wavelength was deduced, using Eq. (13) or its equivalent.

The optical system used at the rocket test stand is illustrated in Fig. 7. The vibration conditions near the rocket flame were so severe that it was necessary to use a periscope tube to transmit the flame light to another place where the interferometer-camera system could operate satisfactorily when protected further by a heavy, air-tight, steel housing box placed on shock mounts. The "2F" lens system of the periscope[1] was arranged to form a real image of the portion of the flame in the field of view on the window of the housing box where a mask could be placed to exclude all light except that which originated in the zone of interest in the flame. A swinging mirror assembly was placed at the flame end of the periscope so that the line of sight could be made to intersect the flame at any desired angle between the limiting upstream and downstream angles at which there was danger of the flame burning the mirror frame.

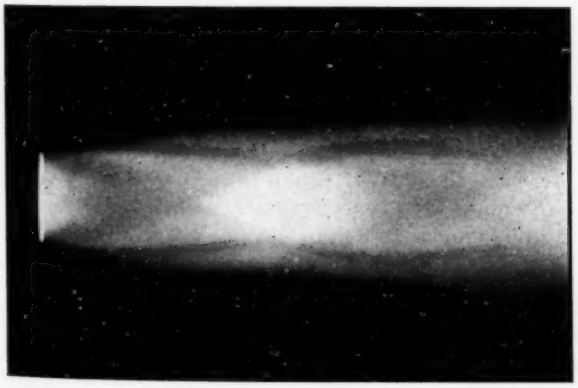


FIG. 8. Photograph of an oxygen-alcohol rocket flame.

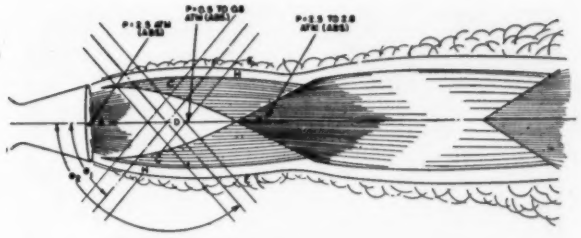


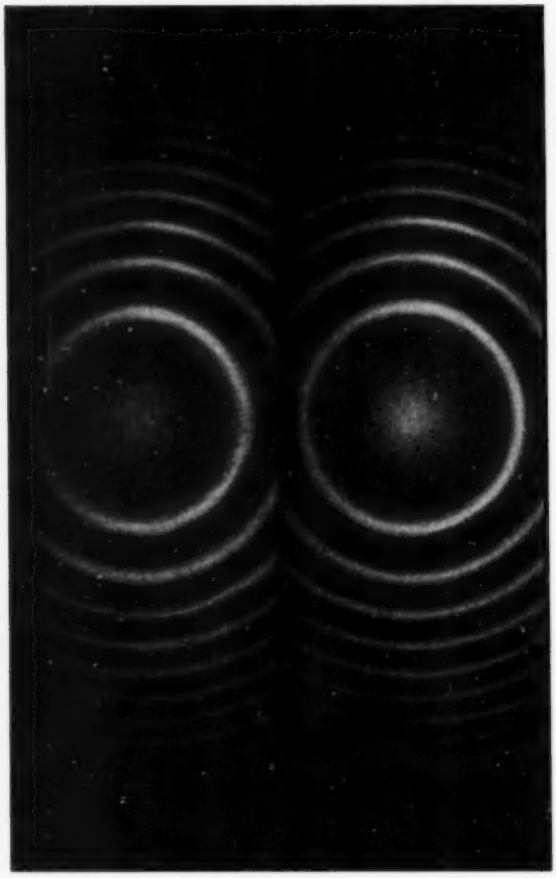
FIG. 9. Diagram of a rocket flame showing preferred lines of sight for velocity measurement by the doppler shift method.

A filter was used in front of the interferometer to exclude as much as possible of all light except the spectral lines being used. For sodium D-line radiation both a Farrand interference-type filter and a Wratten E-22 filter were used on different occasions. The Wratten filter, in combination with 103-T Eastman plates, gave better exposure efficiency with adequate wavelength discrimination. The filter cut off the radiation on the short wavelength side, and the 103-T plate was insensitive on the long wavelength side. For lithium 6707A light, combined Wratten E-22 and 33 filters were used with Eastman 103a-F spectroscopic plates. The camera was equipped with special plate holder and shutter mechanisms which could be operated quickly by remote control in order to get an upstream and a downstream exposure during one rocket motor test run.

IV. MEASUREMENT OF GAS FLOW VELOCITY

To determine the flow velocity of the gas in the rocket exhaust stream it was necessary to obtain two interference patterns, one using light emitted in an upstream direction, and another for light emitted in a downstream direction. On account of the complicated temperature and pressure structure of the flame it was necessary to select lines of sight through the flame which would not include the shock cones or thick regions of cool boundary layer. Figure 8 is photograph of a typical flame. Figure 9 is a diagram taken from the photograph which shows the zones more clearly. The region D is one of over-expansion, where the pressure is below atmospheric, the temperature and luminosity are relatively low, and the gas velocity is

[1] F. P. Bundy and H. M. Strong, J. Opt. Soc. Am. **39**, 393 (1949).



2 1

FIG. 10. Photograph of a pair of doppler interference patterns taken from sodium *D*-line radiation from a rocket flame.

greater than that corresponding to expansion to atmospheric pressure. The region *F* is the shock cone, where the pressure and temperature are both quite high. In region *C* the temperature, pressure and velocity are about what they should be for expansion to atmospheric pressure. It is obvious that if the lines of sight passed through region *F* the results would be thrown in error by pressure shift of wavelength, and, in addition, the spectral line contour would be complicated by the passage of light through so many zones of different temperature. The two lines of sight shown in the drawing were selected as being best for freedom from strong pressure zones, for parallelism of flow to the flame axis, and for relative thinness of boundary layer.

Using the geometry of Fig. 9, the doppler shift relationship is

$$\delta\lambda/\lambda=(v/c)(\cos\theta_1-\cos\theta_2), \qquad (14)$$

where v is the velocity of the flame gas parallel to the flame axis, and c is the velocity of light. The gas velocity in this case is known to be of the order of 2×10^5 cm/sec, while the velocity of light is 3×10^{10} cm/sec, making the ratio $v/c=(2/3)\times10^{-5}$. The attainable experimental values of $(\cos\theta_1-\cos\theta_2)$ are only slightly greater than unity, and hence the ratio of $\delta\lambda/\lambda$ is about 10^{-5}. The mean value of λ for sodium is 5893A, which makes the value of $\delta\lambda$ to be measured about 0.06A. The wavelength interval between successive fringes for the interferometer set at $d=0.2034$ cm is 0.85A. Thus the doppler shift of the fringes to be expected is less than one-tenth of a fringe spacing. To show that this is observed in practice a typical pair of interference patterns is shown in Fig. 10, and the densitometer traces taken across their diameters are shown superposed in Fig. 11.

To get a result accurate to ±5 percent from such a small shift it is necessary that the interferometer-camera equipment be nearly perfect so that the interference patterns themselves can be dependable. The necessity for having narrow spectral lines and hence narrow fringes can be readily appreciated from Fig. 11. One of the most important pieces of equipment required is the recording densitometer for making the density traces across the diameters of the fringe patterns. The machine used for this purpose must have almost perfect interlocking of motion between the scanning carriage bearing the photograph plate and the recorder paper on which the pattern is traced. The authors used a Leeds and Northrup Microdensitometer. Usually the 2-cm span on the diameter of a fringe pattern was expanded to about 50 cm on the densitometer tracing. The two densitometer tracings taken from a pair of interference patterns were then superposed on an opal glass plate illuminated from below for measurement of the change in diameter of the fringes. Measurements on the three or four fringe circles nearest center were usually made. These gave three or four values of $D\delta D$ which could be averaged.

Most of the measurements were made using sodium *D*-line radiation, but on a few occasions the resonance radiation of lithium 6707A was used in order not to interfere with some sodium line reversal temperature measurements which were being made simultaneously and which required a much higher concentration of sodium in the flame than could be tolerated for the velocity measurement.

The results of a series of twelve test runs are given in Table I in which the gas velocities determined by the doppler shift method are compared with those deduced from the mass flow rate and thrust data (specific impulse). The velocities given by the doppler shift method average about 6 percent higher than those from the other method. The variation in the amount of the difference is quite irregular and indicates that the accuracy of any one measurement by the doppler method is not better than ±8 percent.

When the distribution of the mass flow in the different zones of the flame is considered, it is seen that the two methods should not agree exactly. Referring again to Fig. 9, consider the line of sight at θ_1. The flame zones in the line of sight are, in the order of their proximity to the mirror, the thin subsonic boundary

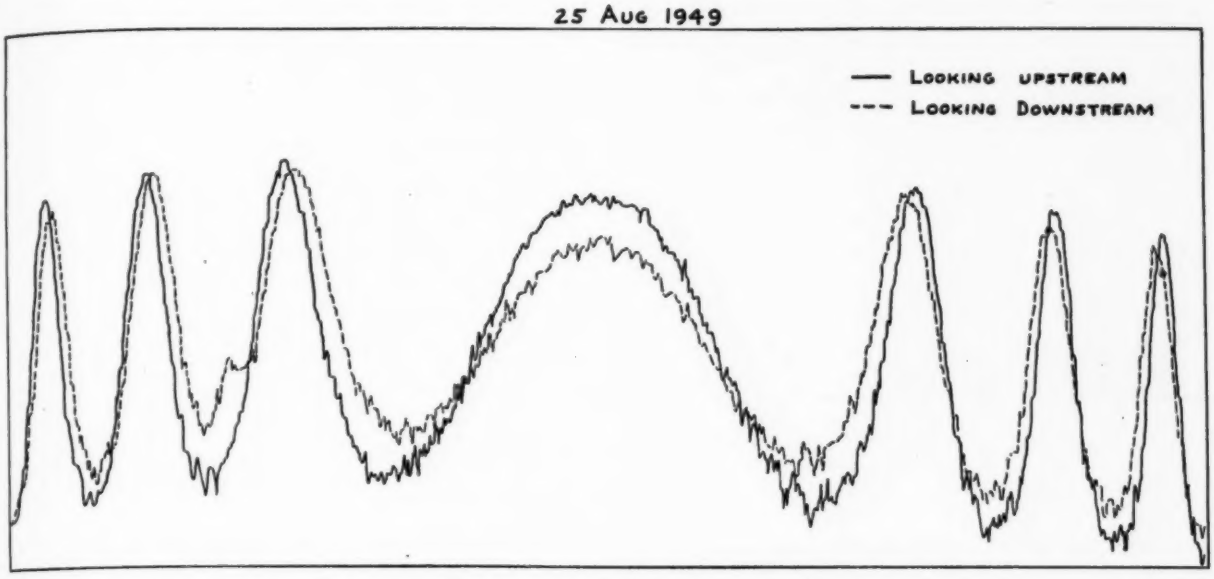


FIG. 11. Densitometer traces across the diameters of the pair of doppler interference patterns shown in Fig. 10. (August 25, 1949.)

layer E, the relatively cool faintly luminous supersonic zone H, the hot supersonic zone C, the relatively cool overexpanded faintly luminous supersonic zone D, then repeated sections of zones C, H, and E. Spectroscopically, the two sections of zone C contribute the most light, zone D considerably less, and zones H and E very little. This is not representative of the mass flow because at least one-fourth of the mass flow is in zone H, less than half in zone C, and about one-third in zone D. On account of over-expansion the gas velocity in zone D may be as high as 2.6×10^5 cm/sec, while owing to the cooling and friction by the walls of the combustion chamber and nozzle the gas in zone H has a velocity lower than average. Estimating the distribution of mass flow and luminosity in the different zones as accurately as possible it is found that the spectroscopic doppler shift method should indicate a gas velocity about 2 percent to 6 percent higher than that from the specific impulse method. The observed average difference of 6 percent is in line with this, and it can be concluded that the two methods are in fair agreement. Neither method is capable of giving the detailed velocities at particular zones in the flame.

As mentioned earlier in this paper there were observed during this study some interesting effects on the spectral line contours arising from difference of doppler shift in different zones of the flame. In Fig. 9 consider the line of sight corresponding to θ_2. The outer zones E and H are considerably cooler and less luminous than the zone C. The boundary layer zone E also has very low velocity. Consequently, the wavelength of maximum emission of zone C will be shifted toward the short wavelength side of the maximum of absorption of zone E. This causes the contour of the sodium lines to become unsymmetrical. Each fringe in the interference pattern shown in Fig. 12 shows this clearly. By contrast, when the line of sight was perpendicular to the flame axis, so that there was no doppler shift, the line contour became symmetrical as shown by each fringe in the interference pattern presented in Fig. 13.

V. MEASUREMENT OF GAS PRESSURE

For the gas pressure measurements the swinging mirror assembly on the periscope tube was set to view the flame perpendicularly in order to eliminate any doppler shift in wavelength of the spectral lines. Light from the region of interest in the flame was allowed to pass into the interferometer-camera assembly through a properly shaped aperture in the mask placed on the window of the box housing the interferometer-camera assembly as described earlier. Again two separate interference patterns had to be taken; one from flame light during a test run of the rocket motor, and a second one from the light of a standard sodium vapor lamp inserted in the same position originally occupied by the flame. Comparison of these two interference patterns gives the wavelength difference between the sodium

TABLE I. Gas velocities in rocket exhaust flame.

Nominal run No.	Spectral line	Velocity (cm/sec) $\times10^{-5}$ Doppler	Specific impulse	% Difference
1	Sodium 5893A	1.96	2.07	−5.3
2	Sodium 5893A	2.26	2.20	2.7
3	Sodium 5893A	2.17	2.17	0.0
4	Sodium 5893A	2.14	2.16	−0.9
5	Sodium 5893A	2.34	2.18	7.3
6	Sodium 5893A	2.28	2.16	5.5
7	Lithium 6707A	2.26	2.11	7.1
8	Lithium 6707A	2.31	2.10	10.0
9	Sodium 5893A	2.30	2.29	0.4
10	Sodium 5893A	2.35	2.19	7.3
11	Sodium 5893A	2.51	2.19	14.6
12	Sodium 5893A	2.39	2.12	12.7

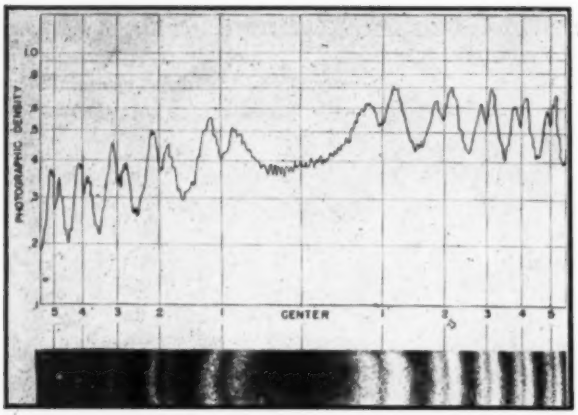


FIG. 12. Unsymmetrical dimpled sodium line intensity contours resulting from difference of velocity and temperature in the inner and outer zones of the flame.

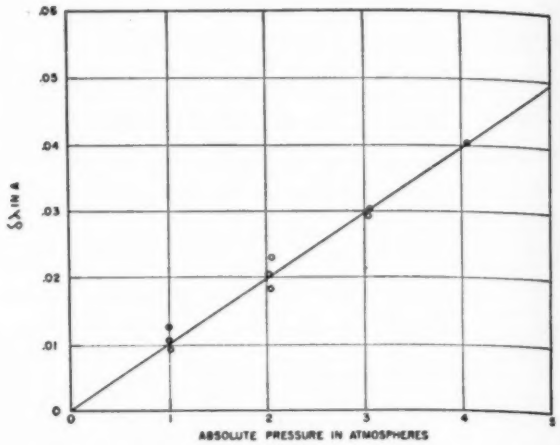


Fig. 14. Redward shift of wavelength of sodium *D*-line radiation with flame gas pressure.

vapor lamp radiation (at practically zero gas pressure) and the sodium radiation from the flame. From the calibration curve for shift of wavelength with pressure shown in Fig. 14, the gas pressure in the line of sight in the flame was determined.

No references to previous measurements of the pressure shift of wavelength of sodium *D* line radiation in flame gases could be found in the literature. A rather large number of measurements had been made for sodium in nonoxidizing gases such as hydrogen, nitrogen, and argon[2] by absorption processes at low temperatures rather than by emission processes at high temperatures. Margenau and Watson found shifts of 0.068A, 0.060A, and 0.051A for A, N_2, and H_2 gases, respectively, per atmosphere pressure at 273°K. These values correspond to 0.021A, 0.018A, and 0.016A per atmosphere at 2600°K. As the rocket flame gas (approximately 56 percent H_2O, 25 percent CO_2, 13 percent CO, 5 percent H_2) is quite different from any of these it was necessary to set up a flame chamber in the laboratory and burn a flame in it at various known pressures to determine the shift of wavelength of the sodium *D* radiation as a function of the absolute gas pressure.

Figure 15 is a photograph of the apparatus which was used for this purpose. Oxygen and propane gases were burned together by a special blast lamp located in the center of the large horizontal cylindrical chamber so that the flame was in the line of sight of the windows in each end. The condensible products of combustion were liquefied in the water-cooled condenser arm, while the remaining gases escaped through a relief valve which was set for the desired chamber pressure. Interference patterns were made using the sodium *D* radiation from the flame and compared with the patterns made at the same time with the same optical setup from the radiation of the standard sodium vapor lamp. The calibration curve obtained is shown in Fig. 14. The wavelength shift is proportional to the absolute pressure, as it should be according to theory, and has a magnitude of about 0.010A per atmosphere at 2600°K.

The results of the pressure measurements made on the gases of the rocket flame are shown in Fig. 9. At zone *A*, just outside the nozzle the pressure was about 2.5

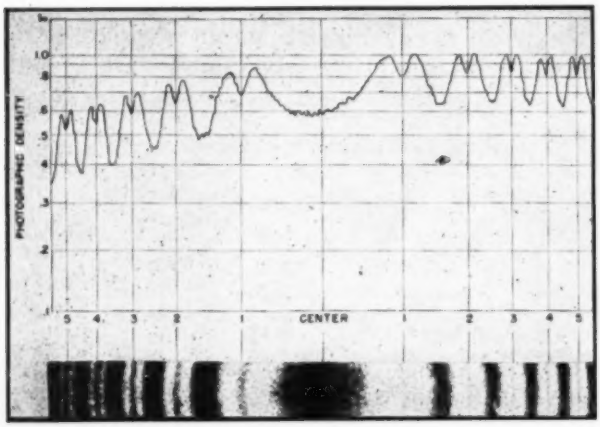


FIG. 13. Symmetrical dimpled sodium line intensity contours from light emitted perpendicular to zoned flame.

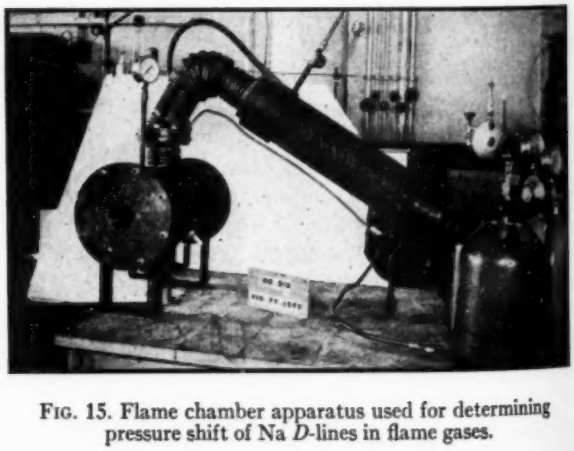


FIG. 15. Flame chamber apparatus used for determining pressure shift of Na *D*-lines in flame gases.

[2] H. Margenau and W. W. Watson, Phys. Rev. 44, 92 (1933).

atmospheres (absolute). In zone D, which is a region of over expansion, the pressure is below atmospheric, roughly 0.5 to 0.8 atmosphere. This could not be determined very accurately because most of the radiation in the measurement came from zone C, and this had to be allowed for. In the front part of the shock cone the pressure was found to be about 2.5 to 2.8 atmospheres. All these pressures are roughly consistent with those to be expected theoretically in a supersonic gas stream of this nature emerging from a diverging, underexpanding nozzle into the atmosphere at normal pressure.

VI. DISCUSSION AND CONCLUSIONS

This investigation has proved that it is possible to measure both velocity and pressure of the gases in rocket exhaust streams with fair accuracy by spectroscopic methods which require no disturbance of the flame by mechanical probes or the introduction of foreign nongaseous matter in the flame. In the measurement of both velocity and pressure the values indicated are a kind of weighted average of the actual values along the line of sight in the flame, if the flame is not homogeneous. The values for particular zones within the flame can only be estimated by considering the flame geometry along with the average value for the line of sight. As the methods require the presence of free sodium atoms they can be applied only to those gas streams which are hot enough to vaporize and decompose sodium compounds.

The absolute accuracy of any one velocity measurement is no better than about $\pm 1.5\times 10^4$ cm/sec, and hence there would be little point in trying to apply the method to gas streams having velocities less than 10^5 cm/sec. Similarly, the accuracy of this method of pressure measurement is of the order of ± 0.3 atmosphere, so that it is applicable only where relatively large pressure changes occur in the gas stream.

The authors wish to express their appreciation to A. J. Nerad, Research Laboratory, who suggested the problem and provided constant encouragement, and to E. H. Hull and his associates at Malta Test Station for providing facilities for the experiments made at the Rocket Test Stands.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1951

Air Condensation in Hypersonic Flow

P. WEGENER, S. REED, JR., E. STOLLENWERK, AND G. LUNDQUIST*
U. S. Naval Ordnance Laboratory, White Oak, Maryland

(Received March 16, 1951)

The NOL 12×12-cm hyperballistics wind tunnel No. 4, which operates in a Mach number range up to eleven, is briefly described. Experiments performed in this tunnel indicate that a fraction of the air condenses at or shortly after reaching the dew line for air condensation unless the supply air is preheated. The measured effects of air condensation on the flow at high Mach numbers for a variety of supply temperatures and pressures are given. It is found that of the flow parameters commonly measured in supersonic wind tunnels, only static-pressure and shock-angle measurements are sensitive to liquefaction effects. A simple thermodynamic analysis assuming a reversible condensation process allows interpretation of some of the observations.

1. INTRODUCTION

ONE of the problems in present-day research on hypervelocity missiles and projectiles is the development of an adequate hypersonic wind tunnel. (The word "hypersonic" is used here to describe flows at Mach numbers higher than about five.) It is understood that the mean free path in the flow is negligibly small with respect to the boundary-layer thickness on a body. To insure the latter condition for expansions to high Mach numbers, hypersonic tunnels must operate at high supply pressures. Then a rough calculation shows that such expansions of dry air beginning at high initial pressure and room temperature lead to pressures and temperatures at high Mach numbers which lie in the coexistence region of air. High wind-tunnel supply temperatures would keep the air above its dew line. On the other hand, from experimental investigations of steam and of moist air expansions below the dew point for water vapor, it was known that the water vapor supersaturates before condensing. Discussions among workers in the field during recent years centered about the applicability of the steam or water vapor experience to the air case.

In order to investigate this and other basic physical problems of high Mach number flow, the Naval Ordnance Laboratory 12×12-cm Hyperballistics Tunnel No. 4 was designed and built. Shakedown operations began in May, 1950; and this paper represents some of the work done to date on the air condensation problem.

2. BRIEF DESCRIPTION OF THE NOL 12 × 12-cm HYPERBALLISTICS TUNNEL

Figure 1 shows a schematic diagram of the wind tunnel and its auxiliary equipment. Filtered and dried air is stored in tanks at 3000-psi pressure. This air sup-

*Members of Hyperballistics Subdivision III, Aeroballistic Research Department.

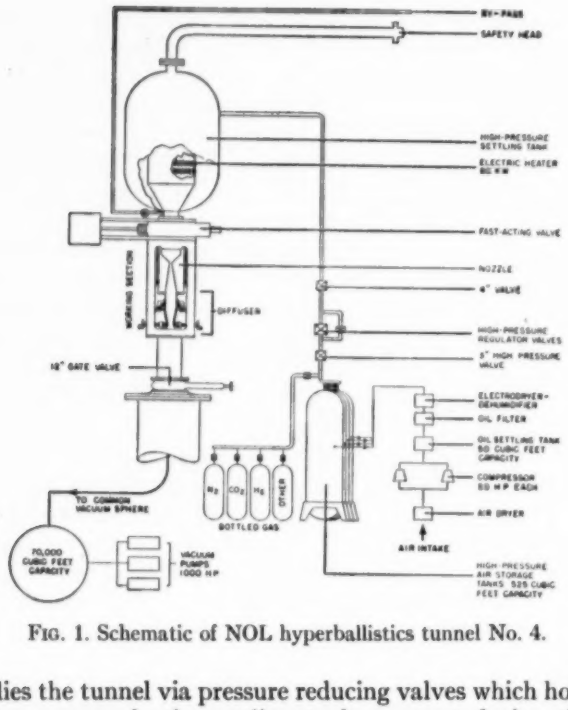


FIG. 1. Schematic of NOL hyperballistics tunnel No. 4.

plies the tunnel via pressure reducing valves which hold the pressure in the settling tank constant during the run. The settling tank contains an 80-kw electric heater. Electronic and mechanical controls hold the air temperature constant at a desired value during the run. So far, supply pressures (p_0) up to 8 atmos and supply temperatures (T_0) up to 300°C have been used. After passing through the vertical working section, the air enters vacuum pumps via a 2000-m³ vacuum sphere. Since air supply and vacuum capacity exceed the demand, the tunnel operates continuously.

For the purpose of investigating the physical properties of the flow under a wide variety of conditions, a slender wedge-type nozzle with a 12×12-cm exit area was employed. The potential-flow Mach number, obtained theoretically from the ratio of exit area to throat of this nozzle, will be referred to as M_s, the Mach-number setting. The useful M_s-range is from 4 to 11. The experimental Mach number in the nozzle is always slightly lower than M_s owing to the presence of the boundary layer. Preheating the air above room temperature causes the nozzle surfaces to become warm. As a consequence the throat gradually closes in the course of a "hot" run. The resulting change of M_s during a sequence of readings was noted and will be indicated in some of the figures which follow.

3. SURVEY OF THE AIR CONDENSATION PROBLEM

It is known that the expansion of air in a wind tunnel follows a perfect-gas isentrope up to the pressure and temperature of condensation. For given supply conditions (denoted by subscript 0) the intersection of the isentrope and the dew line in a p-v diagram can be identified with a Mach number M_c, called the Mach number of equilibrium condensation. For example, if $p_0=1$ atmos and $T_0=15$°C, then from the appropriate isentrope and the dew line one obtains roughly the value $M_c=4.8$. A study of the isentrope and dew line $p-T$ diagram shows that M_c can be increased by means of decreasing p_0 or increasing T_0. High supply pressures are mandatory at high Mach numbers in order to maintain a sufficiently large density in the test section so that slip-flow effects do not occur. Therefore, only increasing T_0 is a feasible means of increasing M_c. But high values of T_0 are needed to attain high values of M_c; for example, to obtain $M_c=10$ with $p_0=100$ atmos, T_0 must be about 800°C.

There was some expectation during the last years that the air would undergo large supersaturation in the wind-tunnel expansion. This seemed to be verified by tests performed with a small wind tunnel at Princeton.[1] Moreover, the experience with water-vapor expansions pointed to this conclusion. It was known that water vapor, expanding as steam in turbine nozzles, or expanding together with air in supersonic wind tunnels, followed a "dry isentrope" well beyond saturation. When condensation of water vapor occurs in supersonic wind tunnels, it is a sudden process which gives rise to a so-called condensation shock. In tests performed at this Laboratory with a 2.5×2.5-cm wind tunnel, water vapor (in air) was found to undergo a dry isentropic expansion 80 to 100 degrees Kelvin beyond its dew point.[2]

It was also known that water vapor supersaturated when expanded quickly in cloud chambers. Some results of Cwilong[3] are of particular interest in this respect. He found that if moist air, from which foreign nuclei were removed by filtering, was expanded rapidly in a cloud chamber below 150°K, condensation could not be observed. Head[4] found the same phenomenon in a supersonic wind tunnel. However, two of the authors, using a sensitive photomultiplier tube for detecting light scattering, found that if moist air was expanded very quickly in a shock tube and observed after a 4-millisecond period in which it remained at about 130°K, it definitely contained droplets or crystals.[5]

A successful theoretical treatment of the water-vapor condensation shocks in nozzle flow was given by Oswatitsch,[6] using the condensation theory of Becker, Doering, and Volmer.[7] This theory treats the equilibrium condensation of a vapor in which foreign nuclei are absent. Calculations for air condensation, using this same theory, predict a considerable degree of supersaturation in the wind-tunnel expansion.

These expectations of large air supersaturations were not fulfilled in our tests. Rather, our results agree with

[1] S. M. Bogdonoff and L. Lees, Department of Aeronautical Engineering, Princeton University Report No. 146 (1949).
[2] P. Wegener, Phys. Rev. **76**, 6 (1949).
[3] B. M. Cwilong, Proc. Roy. Soc. (London) **A190**, 137 (1947).
[4] R. Head, GALCIT Dissertation (1949).
[5] P. Wegener and G. Lundquist, J. Appl. Phys. **22**, 2 (1951).
[6] K. Oswatitsch, *Jahrb. Deutsch. Luftf. Forsch.* (1941).
[7] M. Volmer, *Kinetik der Phasenbildung* (Steinkopff, Dresden, 1939).

the data found at the NACA Langley Aeronautical Laboratory by C. H. McLellan, which indicate that air condensation occurs at or shortly after reaching the calculated dew line.[8] Our results at low temperatures agree also with those found with the California Institute of Technology hypersonic tunnel by Nagamatsu and his group.[9]

A plausible explanation of this absence of considerable air supersaturation, as would be expected according to the theory of Becker, *et al.*, is that foreign nuclei which are larger than the critical size[7] are present. An earlier condensation of a small fraction of the CO_2 or of the small amount of unavoidable water vapor (here about 2 parts or less in 100,000 by weight) present in the wind-tunnel air would supply a sufficient quantity of such nuclei.

4. NOL AIR CONDENSATION RESULTS

Tests were performed in the M_s range from 6 to 9, and all the measurements showed a consistent behavior. For this reason an account of results for only one representative Mach number ($M_s=7.6$) is presented in this report. Measurements shown will be of static pressure, pitot pressure, and shock-wave angles.

(Light-scattering experiments show the presence of droplets or crystals in the test section at low supply temperatures and high Mach numbers. The light scattering disappears if the incident light beam is moved into the boundary layer or behind strong shock waves. The scattered light also disappears if the air is preheated. Since qualitative and quantitative evaluation of such data is difficult, the measurements will not be exhibited.)

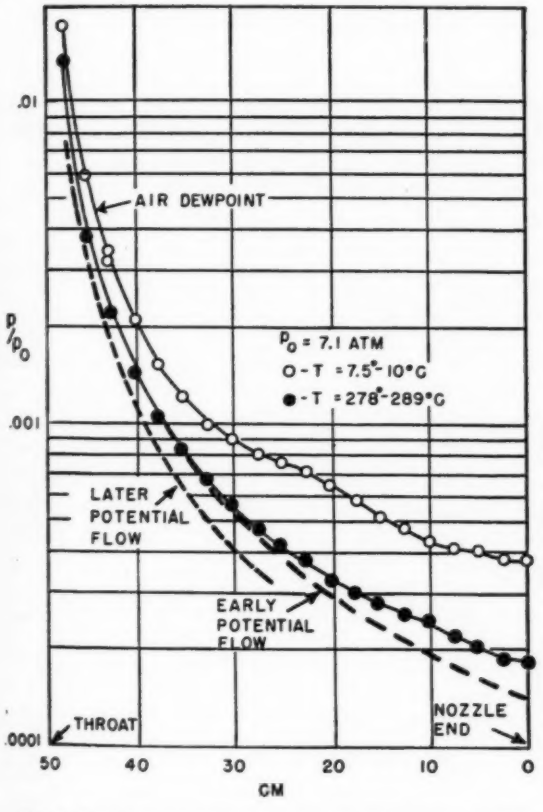


FIG. 2. Static pressure measurements on the nozzle wall center line at $M_s=7.6$.

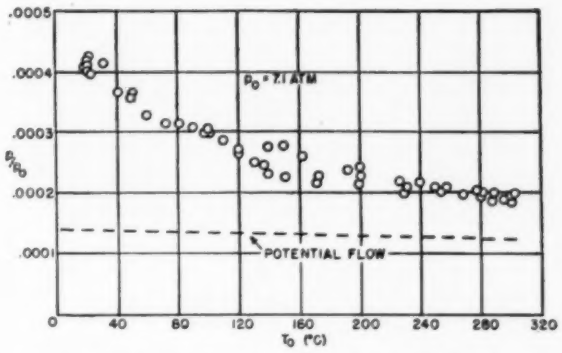
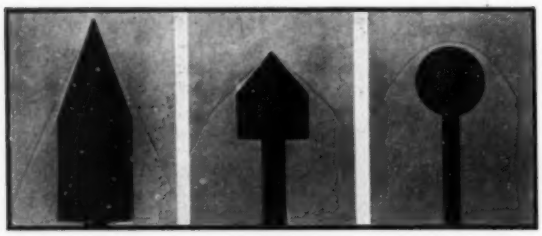


FIG. 3. Static pressure measurement at nozzle exit as a function of supply temperature at $M_s=7.6$.

(a) *Static pressure measurements* were made along the center line of the nozzle walls. Results for $M_s=7.6$ are shown in Fig. 2. The open circles were obtained with T_0 about room temperature, called a cold run, and the solid circles with T_0 raised so that $M_s=M_c$, called a hot run. The static pressures expected for a nonviscous isentropic expansion of a perfect gas are shown as dotted curves on each figure. Figure 2 exhibits two such potential-flow curves, the one for $M_s=7.6$ corresponding to a cold throat, the other for the hot throat. The hot-throat, potential-flow curve refers to measurements made at the end of the run (called "later potential flow") and corresponds to $M_s=7.9$. In this case, the data taking began at the nozzle exit and proceeded towards the throat. The difference between the hot-run dimensionless static pressure and that of the potential flow may be attributed in large part to the boundary layer which reduces the effective area. The difference between the cold-run and hot-run pressure values began to show up shortly after the predicted saturation point of the air and was taken as evidence for the presence of condensation. Figure 3 shows data on the dimensionless static pressure, measured at the nozzle end as function of supply temperature T_0, for $M_s=7.6$. Between the extreme T_0 values, the static pressure changes by a factor of two. The pressure data in Fig. 3 are measurements made at the nozzle wall or on the nozzle center line, with silicone oil, butyl phthalate, or McLeod manometers, for runs during which T_0 increased and for runs during which T_0 decreased.

(b) Shadowgraphs obtained when a 40° and 80° cone-cylinder body and a sphere are placed in the test section at $M_s=7.6$, with T_0 approximately 270°C, are shown in Figs. 4, 5, and 6. Since no condensation is present at this T_0, the actual Mach number is 7.2 for these three pictures. The evaluation of the shock

[8] J. T. Becker, J. Appl. Phys. **21**, 7 (1950).
[9] H. J. Nagamatsu, Phys. Rev. **79**, 546(T) (1950).



FIGS. 4–6. Shadowgraphs (reading from left to right) of 40° cone-cylinder, 80° cone-cylinder, and sphere at $M=7.2$ ($p_0=7.1$ atmos, $T_0=270°C$).

angles on the two cones by independent observers for three values of T_0 is shown in Fig. 7. The shock angle on the 40° cone decreased as T_0 increased. The shock angle on the 80° cone, and also for all strong or near-normal shocks we observed, showed no systematic variation with changing T_0. (The M-scales in Fig. 7 do not take the boundary layer on the cones into account. Therefore, the Mach number from shock angle even at high T_0 is lower than the actual flow Mach number.)

(c) Figure 8 shows the dimensionless *Pitot pressure* (p_0'/p_0) measured at the center of the nozzle exit as a function of supply temperature T_0. Evidently p_0'/p_0 is rather insensitive to changes in T_0 throughout the T_0 range where condensation effects may be expected.

Isentropic flow tables can be used formally for the conversion of pressure measurements into values of Mach number. This procedure, which ignores condensation effects, gives the results shown in Fig. 9 in which the nozzle exit data for $M_s=7.6$ were used. From Fig. 9

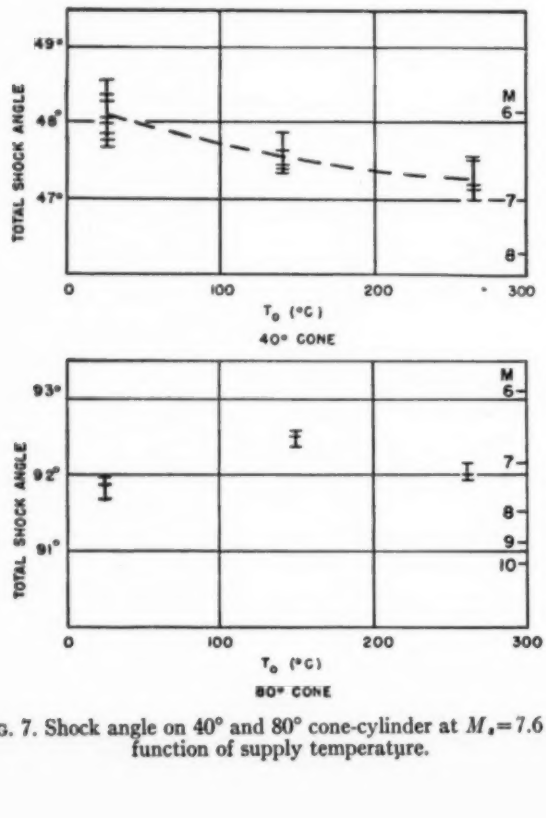


FIG. 7. Shock angle on 40° and 80° cone-cylinder at $M_s=7.6$ as a function of supply temperature.

it can be inferred that these formal Mach numbers agree with one another for supply temperatures T_0 less than the value T_{0c}, for which $M_c=M_s$. If this agreement in Mach number can be taken as a criterion for the establishment of nearly-perfect-gas isentropic flow conditions, it suggests that lower supply temperatures than T_{0c} may be sufficient for many testing purposes in our tunnel. In view of the technical difficulties entailed in achieving high values of T_0, this conclusion may be of practical importance. At present three explanations may be advanced for this apparent saving in supply temperature: (i) Since the dew line (Fig. 9) is determined from extrapolated data, its location is uncertain. Therefore, a definite minimum supply temperature to avoid condensation is yet unknown. (ii) A small amount of liquefied air may go undetected owing to the inaccuracy of static pressure and pitot-pressure measurements, from which M is determined. (See scatter of measurements in Figs. 3 and 8.) (iii) Finally, it is probable that some supersaturation occurs in the initial stages of the expansion. The exact location of deviation of the hot-run and cold-run expansions (Fig.

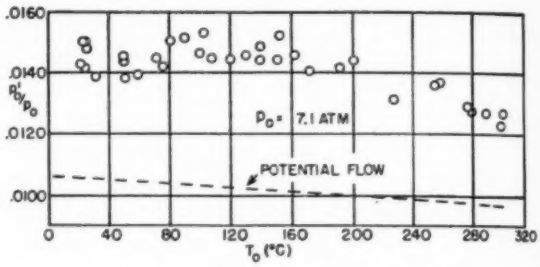


FIG. 8. Pitot pressure measurement at nozzle exit as a function of supply temperature at $M_s=7.6$.

2) is hard to determine experimentally. Light-scattering experiments by Stever[10] appear to indicate some supersaturation. He observed the onset of light scattering downstream of the theoretically predicted saturation point but not at it. It is to be expected further that air with different amounts of impurities might reach different degrees of supersaturation. In fact, there is no reason to believe that "pure" nitrogen expansions would not show supersaturations comparable to those observed for water vapor. Achieving such purity on a technical scale, however, is a different matter.

5. THERMODYNAMICS OF CONDENSATION

The experimental data suggest a simplified model of the condensation process, neglecting supersaturation, in which it is supposed that all processes, except those involving shocks, are reversible.[11] Figures 10 and 11 show features of an isentropic expansion of a single component substance which ends in the coexistence region

[10] H. G. Stever and K. C. Rathbun, Final Report NACA Contract NAW 5713, M.I.T., Aer. Eng. Dept. (1950).

[11] A. Stodola, *Steam and Gas Turbines* (McGraw-Hill Book Company, Inc., New York, 1927).

(shaded).[12] Figure 10 shows a schematic p-v diagram. One expansion, ODE, proceeds as a perfect-gas isentrope from conditions at point 0 to the dew-point line CC', thence as the isentropic expansion of a saturated vapor and its coexistent condensed phase to the end point E. The other expansion, ODE', follows the continuation of the perfect gas isentrope, as it might take place if supersaturation occurs, to the end point E'. The fraction of mass condensed g at the point along ODE is given by the ratio of lengths JB'/BB' along the isotherm BB'. As long as the expansion begins at volumes (specific) much larger than the critical volume, we see from Fig. 10 that g, although continuously increasing, will always remain small. Pressure and temperature along the saturated isentrope DE must continually decrease because the isentrope through a point must always be steeper than the isotherm through the same point.

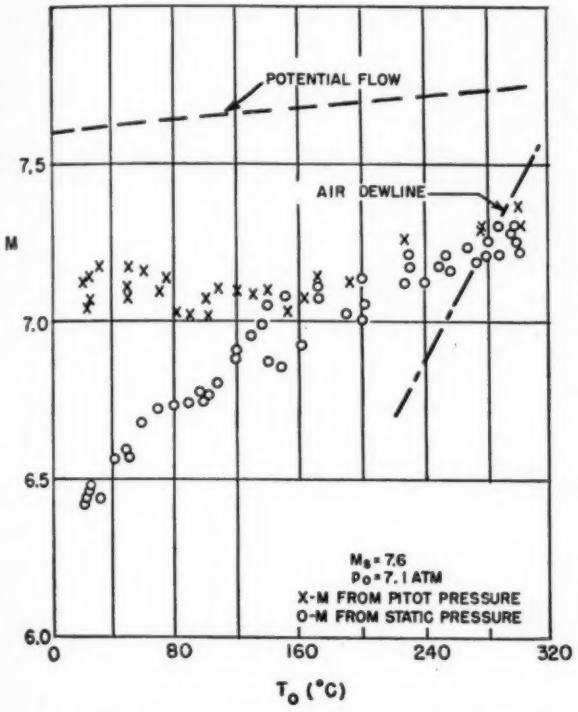


FIG. 9. Mach number evaluated with isentropic flow tables at $M_s = 7.6$ as a function of supply temperature.

An increase in temperature of the starting point 0 accomplishes a displacement of the curve ODE to the right in Fig. 10 and moves the points E, E', and D nearer one another. For the starting temperature T_{0c} at which E, E', and D coincide and for starting temperatures $> T_{0c}$, no condensation is expected.

The schematic Fig. 11 shows slopes $(\partial p/\partial \rho)_s$ taken along the curve ODE in Fig. 10 (full line), and along ODE' (dotted line beyond D). At the point D, $(\partial p/\partial \rho)_s$ falls sharply; and as the expansion proceeds, the curves

[12] F. Bosnjakovic, *Technische Thermodynamik I* (Theodor Steinkopff Verlag., Leipzig, 1944).

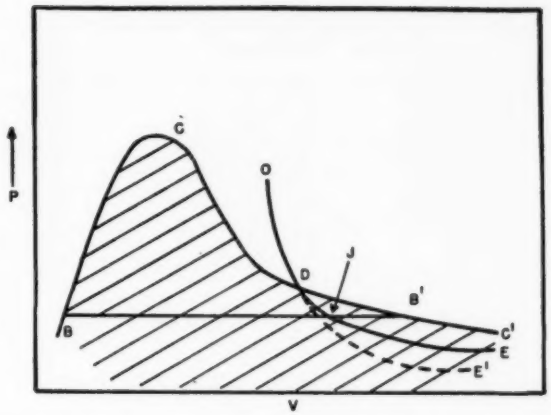


FIG. 10. Schematic pressure-volume diagram.

DE and DE' in Fig. 11 can intersect. If sound speed, after condensation sets in, is given by

$$a = (\partial p/\partial \rho)_s^{\frac{1}{2}}, \tag{1}$$

the curve ODE, Fig. 11, represents a plot of a^2 *versus* volume.

Knowing the location of the point D, we can calculate the mass fraction condensed, g, by the following procedure. The vapor alone is supposed to be a perfect gas so that

$$p = \rho_v RT, \tag{2}$$

where ρ_v is the vapor density and R the gas constant per unit mass of the vapor. For the average over-all density ρ we have

$$1/\rho = [(1-g)/\rho_v] + g/\rho_c,$$

where ρ_c is the density of the condensed phase and can be considered to be very large compared with ρ_v. Hence, we have $\rho(1-g) \cong \rho_v$ and Eq. (2) becomes

$$p = \rho(1-g)RT. \tag{3}$$

The temperature is determined by the pressure alone according to the vapor-pressure equation,

$$T = \varphi(p). \tag{4}$$

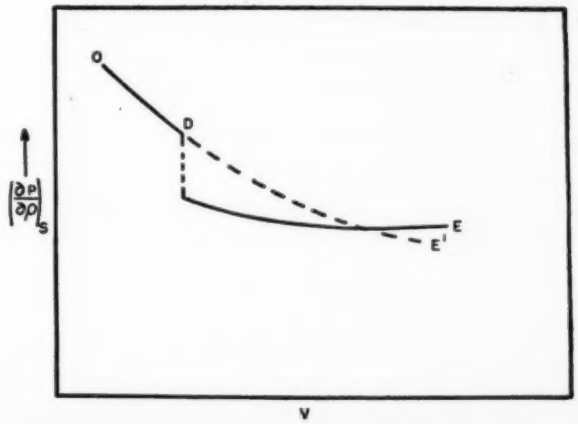


FIG. 11. Slopes of isentropes from schematic p-v diagram.

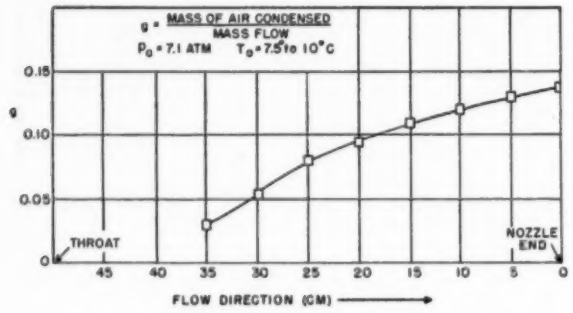


FIG. 12. Condensed mass fraction computed from measurements at $M_e = 7.6$.

For an isentropic expansion, in the coexistence region the (constant) entropy S (per unit mass) can be written

$$S = S_v - gL/T, \tag{5}$$

where S_v = entropy unit mass of the saturated vapor, g = mass condensed/total mass, L = latent heat, a function of temperature. If the value of S is known, g can be calculated from Eqs. (4) and (5). S is known if the point D can be fixed which depends on the starting point 0 and on the location of the dew line CC'.

In order to relate the thermodynamical quantities to flow quantities, we assume that the expansion in the tunnel is isentropic. Then the static-pressure data are that of an expansion similar to ODE in Fig. 10. Knowing p as function of length x along the nozzle center line, measured from the throat, we can find g and ρ from Eqs. (5) and (3). The sound speed, a, is evaluated as a function of x from a plot of measured p *versus* computed ρ. If the value of average mass flow speed v is known at point x, the speed at other points can be found by integrating the equation of motion. One finds

$$v(x) = \left\{ [v(x_1)]^2 - 2 \int_{x_1}^{x} [(dp/dx)/\rho] dx \right\}^{\frac{1}{2}}. \tag{6}$$

Finally, Mach number,

$$M = v/a, \tag{7}$$

can be found for each point along the nozzle.

Knowing ρ and v, we can compute the cross-sectional area A by use of the continuity equation,

$$\rho v A = \text{const.} \tag{8}$$

Conversely, were A known as a function of x and not of p, all necessary information could be obtained. The enthalpy per unit mass h is given by

$$h = h_v - gL = C p_v T - gL, \tag{9}$$

where h_v is the enthalpy of the vapor phase, Cp_v the (assumed constant) specific heat of the vapor, and L the latent heat of vaporization. With decreasing T, the vapor condenses, first forming a liquid, then a solid phase. L may be expected to remain nearly constant at the low temperatures of interest here, except for phase changes.

If vapor and the condensate have the same flow speed, the energy equation is

$$v dv + dh = 0. \tag{10}$$

6. APPLICATION OF THERMODYNAMICS OF CONDENSATION TO EXPERIMENTAL DATA

We can now make some simple numerical estimates of quantities of interest. It is first found by use of Eqs. (6), (8), and the experimental static-pressure data, that the ratio v/a_0 of mass-average flow speed v to sound speed at supply conditions, a_0, is constant to within a few percent for the range of T_0 value of our experiments. This fact implies that the ratio h/h_0 of h to enthalpy h_0 under starting conditions is also nearly constant in the same range of T_0. It implies also that the ratio ρ/ρ_0 at any point x has values for different starting temperatures which are inversely proportional to the respective cross-sectional areas.

To go further, the fact that air is not a single-component substance must be taken into consideration. One can, of course, make rough estimates by neglecting this complication and calculate with extrapolated equations which fit available experimental data on vapor pressure and latent heat of air.[13] Most of the numerical results reported here were obtained by essentially this method and are, therefore, subject to its inaccuracies. Another procedure was tried in which available vapor pressure data for nitrogen and oxygen[14] were used with ideal mixture theory to calculate the dew line for air. The principal difficulty with this procedure is that condensed air is expected to solidify somewhere in the region of temperatures and pressures of interest to us. The frost line for air can be estimated only roughly; its precise location awaits experimental determination. One can say, however, that if the frost line is reached in an expansion, this means that M_e calculated by extrapolating equations based on liquid data will be somewhat too small.

The condensed mass-fraction g was computed by two methods with fairly consistent results. In the first method, the area ratio obtained from the measured pressures for hot-run expansions, which were assumed isentropic, were used as first approximation for area ratios for the cold expansion. The values of ρ/ρ_0 to this approximation are independent of T_0. Results for g found this way are shown in Fig. 12. The other more direct method was to use Eq. (5) with the static pressure and the extrapolated vapor pressures. The values of g obtained in this direct manner turned out to be fairly close to those in Fig. 12. By use of this second method, A can be calculated with an accuracy depending again on the accuracy of the vapor-pressure data. The con-

[13] B. F. Dodge, and A. K. Dunbar, J. Am. Chem. Soc. **49**, 591 (1927).

[14] *The NBS/NACA Tables of Thermal Properties of Gases* (U. S. Government Printing Office, 1950).

sistency of these values of A was compared with that obtained from the hot-run isentrope. The agreement of these two values of A was rather poor. Consequently, agreement of densities, computed by use of the first method with those found using the second method, was equally poor. Graphs of measured pressure p *versus* calculated ρ were drawn, and their slopes evaluated as function of x. According to our assumption, these slopes give

$$(\partial p/\partial \rho)_s = a^2;$$

hence, a Mach number can be evaluated as a function of x. The values of M thus obtained are exhibited in Fig. 13 together with the Mach number obtained formally from isentropic-flow tables by use of the same pressure measurements for the cold run. The hot-run results, where no condensation is expected in the nozzle, are also shown. Of the formally obtained Mach numbers, only the hot-run one is reliable. The variations of the calculated M reflect the uncertainty of the calculation. It is clear, nevertheless, that the calculated Mach number exceeds the others in the early stages of condensation, where the sound speed a falls sharply (see Fig. 11).

The data in Figs. 8 and 9 indicate that Pitot pressure is insensitive to liquefaction effects.

The change in entropy of a perfect gas from the stagnation conditions ahead of the Pitot tube (p_0, T_0) to the conditions in the tube (p_0', T_0), is

$$\Delta S = -R \ln(p_0'/p_0).$$

The Pitot pressure p_0', measured as a function of T_0 with constant p_0, will register changes in ΔS which can be called the "net irreversibility." As long as we deal with a perfect gas under the initial and final conditions regardless of condensation in the course of the process, ΔS is given by the same formula. The results shown in Fig. 3 then can be interpreted to mean that there is no marked change in "net irreversibility" due to condensation. On retention of the assumption that the condensation is reversible, ΔS would be due altogether to the processes occurring in the shock in front of the Pitot tube. The presence of condensation could then alter ΔS only by contributing via its irreversible evaporation in the shock.[4]

7. SUMMARY

The experimental results indicate that within the range of our tests the effects of air liquefaction are the following:

(1) At or shortly after reaching saturation temperature and pressure in the nozzle, the static pressure deviates from that expected of perfect gas behavior.

(2) Different methods of determining Mach number from experimental data yield the same value of Mach number for supply temperatures below that temperature for which no condensation is predicted in the nozzle.

(3) Static-pressure measurements and shock-angle measurements on slender bodies are sensitive to condensation in the flow. Pitot-pressure measurements are not sensitive to condensation.

By use of theory with these results, it is found that:

(4) Within the accuracy of the experiments and available thermodynamic data for air, the deviation of the measured static pressure from the pressure calculated from perfect-gas behavior is as expected for an isentropic expansion of a saturated vapor and its coexistent condensed phase.

(5) The quantity condensed at the nozzle end is never more than a small fraction of the total mass.

(6) The ratio of (mass-average) flow speed to supply sound speed is changed by at most a few percent, due to liquefaction. Also, if the condensation occurs reversibly, the ratio of (mass-average) enthalpy to supply enthalpy is changed only a few percent by liquefaction.

(7) After condensation commences, the temperature of the flow does not decrease as in the condensation free flow.

(8) The ratio of Pitot pressure to supply pressure, as a measure of "net irreversibility" is insensitive to a reversible condensation process.

(9) If liquefaction is present in a nozzle, use of isentropic flow tables with experimental data gives incorrect values of Mach number.

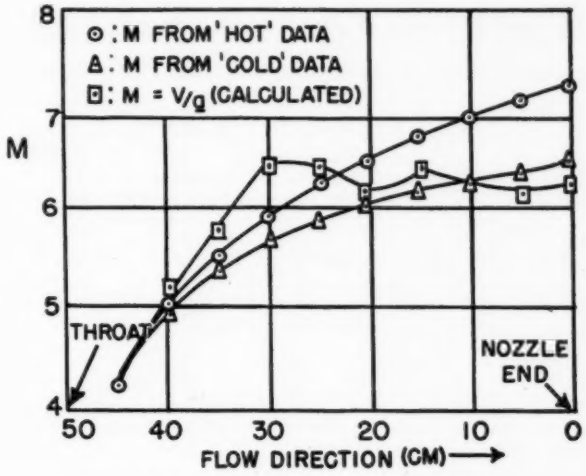


FIG. 13. Mach number determinations as a function of length at $M_s = 7.6$.

ACKNOWLEDGMENTS

The authors wish to thank Miss D. Ericson and Mr. H. Staab of NOL for their cooperation. Numerous discussions with co-workers, in particular with the groups under Dr. H. Nagamatsu at C.I.T., Dr. L. Lees at Princeton University, Dr. H. G. Stever at M.I.T., and Mr. R. Smelt of Aro, Inc., stimulated our work.

Letters to the Editor

On a Theorem of Brouwer

H. F. MATHIS
University of Oklahoma, Norman, Oklahoma
(Received May 28, 1951)

BROUWER[1] has proved that it is impossible to find a vector function **A** which is defined on the bounding surface S of a finite 3-dimensional simply connected region with the property that (a) **A** is continuous everywhere on S, (b) **A** is everywhere tangent to S, and (c) $0<|\mathbf{A}|<\infty$ everywhere on S.

Some interesting applications can be made of this theorem immediately. It follows that in any 3-dimensional flow of fluid around a simply connected solid there must be at least one stagnation point on the surface of the solid. Mathis[2] has used this theorem to prove that an isotropic antenna for electromagnetic radiation or reception is impossible.

A simple corollary is that it is impossible to find continuous functions $f(x, y, z)$, $g(x, y, z)$, and $h(x, y, z)$ such that $xf+yg+zh=0$ and $0<f^2+g^2+h^2<\infty$ for all values of x, y, and z in the region $x^2+y^2+z^2\leq k$, where $k>0$. Consider the dc circuit shown in Fig. 1.

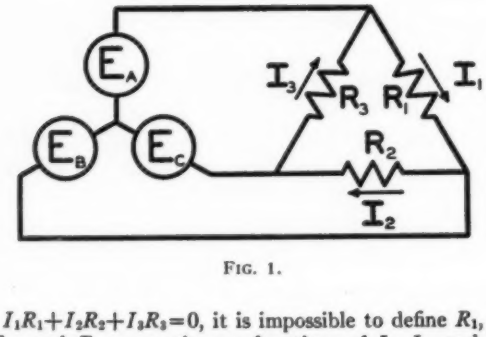


FIG. 1.

Since $I_1R_1+I_2R_2+I_3R_3=0$, it is impossible to define R_1, R_2, R_3, E_A, E_B, and E_C as continuous functions of I_1, I_2, and I_3 with $0<R_1+R_2+R_3<\infty$ for all values of I_1, I_2, and I_3 in the region $I_1^2+I_2^2+I_3^2\leq k$.

[1] L. E. J. Brouwer, Proc. Roy. Acad. (Amsterdam) 11, 850–858 (1909), (English edition) or *ibid.*, 896–904 (Dutch edition).
[2] H. F. Mathis, "A short proof that an isotropic antenna is impossible," to appear in the Proc. Inst. Radio Engrs.

Internal Friction of Cold-Worked Single Crystals of Copper

RYUKITI ROBERT HASIGUTI AND TADAMASA HIRAI
Department of Metallurgy, Faculty of Engineering, University of Tokyo, Tokyo, Japan
(Received April 25, 1951)

IT has been suggested by J. S. Koehler and T. H. Blewitt[1] that the nature of work-hardening of metal crystals would change at small strains, say, in the range from 1 percent to 20 percent. Another evidence for this change could be found in the internal friction *versus* degree of cold-working curves.

We measured the internal friction of single crystals of copper as a function of degree of cold-working, and compared these measurements with stress-strain curves. Specimens of single crystals were made from electrolytic copper by Bridgman's method. The measurements of internal friction were performed by a modified Zener[3] apparatus which has as a detector a condenser formed between a plate and a flat end of the rod specimen vibrating longitudinally. The frequencies of vibration used were from 20 to 100 kilocycles per second.

The results are shown in Fig. 1. The degree of cold-working is expressed in terms of resolved shear strain calculated in the slip direction in the slip plane. The internal friction of crystals annealed at 800°C for two days was from 2×10^{-5} to 2×10^{-4}. It increases rapidly with increasing degree of cold-working, making respective maxima at the shear strains between 0.05 and 0.2 which correspond to the elongations of 1 to 10 percent. These maximum values of internal friction spread between 3×10^{-3} and 10^{-2} according to individual specimens. After the maxima it decreases rapidly as the strain increases, until it comes to a fairly constant value, which is well below the maxima, but above the initial values. No relation has been found until now between the formation of maxima and the directions of crystals with respect to specimen axes. It was confirmed that the occurrence of double slip has nothing to do with these maxima.

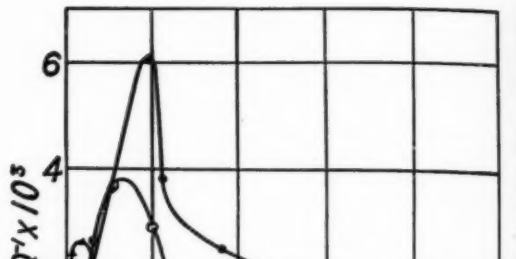


FIG. 1.

Although the amount of resolved shear strain, at which the maximum of internal friction occurs, differs in each specimen, it is always within a narrow range of strain, where the tangent of stress-strain curve increases considerably, as is seen in Fig. 2. In other words, the internal friction begins to decrease when the rate of hardening begins to increase.

The internal friction decreases considerably when the specimen is left at room temperatures and almost resumes its original low value when the crystal is annealed at comparatively low temperatures, say below 150°C, where no recrystallization takes place. And this is true, whatever the degree of cold-working before annealing may be.

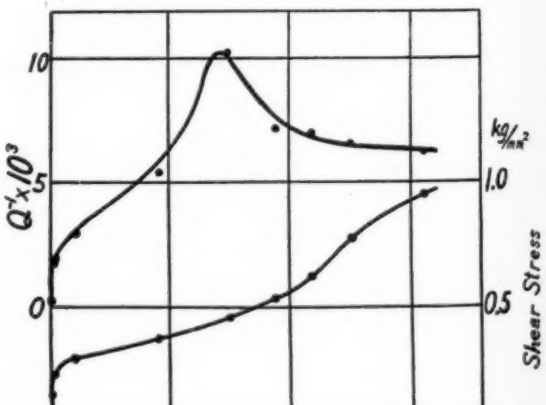


FIG. 2.

A. W. Lawson[3] measured the internal friction of polycrystalline copper as a function of applied stress, and got similar maxima of internal friction. He interpreted the initial increase of internal friction as due to the formation of dislocations which dissipate the energy of vibration, and the following decrease of internal friction as due to the locking of dislocations. In addition to Lawson's suggestion, we should like to suggest the following picture of the behavior of dislocations in connection with our experimental results. The decrease of internal friction at higher strains might be interpreted as due to the gradual increase of the degree of locking of all dislocations. This, however, may not be the case. As the deformation proceeds, most of the dislocations together with newly formed ones will be locked so as not to contribute to the internal friction any more, while some of them will still remain relatively free to move to contribute to the internal friction. The decrease of internal friction at higher strain should be accounted for as due to the decrease of the number of these free dislocations. These relatively free dislocations may easily disappear at low annealing temperatures, but the locked ones may not, in accordance with our experimental results that the internal friction of cold-worked crystals almost resume their original low values at low annealing temperatures, while the hardness of crystals is kept essentially constant. An alternative picture could be proposed for the decrease of internal friction at low annealing temperatures. The dislocations, which we referred to as "relatively free dislocations," might form on annealing a more stable array, something like the one postulated in the dislocation theory of polygonization.[4] Finally, it may be concluded from our results that the increase of relatively free dislocations takes place in the first stage of stress-strain curve, where work-hardening is very little, and that the locking of dislocations predominates in the second stage of stress-strain curve, where the hardening of crystals is considerable.

[1] J. S. Koehler and T. H. Blewitt, Phys. Rev. **75**, 1952 (1949).
[2] C. Zener, Proc. Phys. Soc. (London) **52**, 152 (1940); Randall, Rose, and Zener, Phys. Rev. **56**, 343 (1939).
[3] A. W. Lawson, Phys. Rev. **60**, 15 (1941).
[4] R. W. Cahn, J. Inst. Metals **76**, 121 (1949).

Volume-Temperature Relationships for the Room Temperature Transition in Teflon

FRED A. QUINN, JR., DONALD E. ROBERTS, AND R. N. WORK
National Bureau of Standards, Washington, D. C.
(Received May 4, 1951)

VOLUME–TEMPERATURE relationships have been observed on a sample of Teflon (polytetrafluoroethylene) over the temperature range −35°C to +130°C. The purpose of this work was to investigate the room temperature transition in Teflon and determine thermal expansivities above and below this transition. The room temperature transition has been reported by Rigby and Bunn[1] and was discovered independently by Weir[2] in early 1949 using an interferometric dilatometer in the course of a continuing program of volume-temperature studies of transitions in high polymers. Rigby and Bunn[1] conclude from x-ray studies that the transition is associated with a change in crystal structure. Melting of the crystals has been reported to occur near 325°C.[3,4]

Specific volume is plotted against temperature in Fig. 1. The values were obtained by means of a volume dilatometer using techniques described by Bekkedahl.[5] Mercury was used as the confining liquid. The specimen was cylindrical, about 20 mm in diameter, 45 mm in length, and weighed about 25 grams. The density, determined by hydrostatic weighing, was found to be 2.218 g/cm³ at 25°C. Rigby and Bunn found that the densities of different samples of Teflon differed by as much as 2 percent. Points were obtained from −35°C to +130°C upon increasing the temperature of the specimen at the uniform rate of one degree in five minutes, and from +40°C down to 0°C upon decreasing temperature at the same rate.

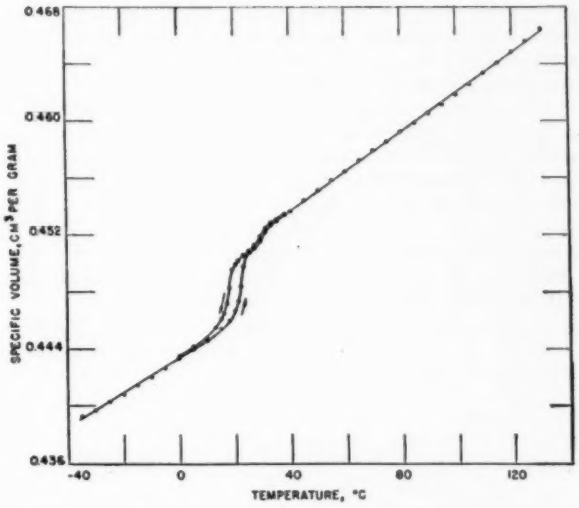


FIG. 1. Specific volume *vs* temperature for Teflon.

It will be observed that the room temperature transition is a double transition. The low temperature portion, which accounts for 85 percent of the volume change in the transition, is centered around +20°C, while the high temperature portion is centered around +30°C. The double nature of this transition has not been reported previously. The slope of the straight portion of the curve above the transition region is 1.37×10^{-4} cm³/g deg C, and below the transition, the slope is 1.17×10^{-4} cm³/g deg C. The straight portions of the curve above and below the transition, extrapolated to 20°C, show a volume change due to the transition of 1.23 percent.

Renfrew and Lewis[3] found a double break in the temperature time curves during formation and melting of crystals near 325°C and have suggested the possibility of two crystalline phases of slightly different melting points.

Figure 2 shows the lower transition region in greater detail and indicates the nonequilibrium character of the transition. Volume-temperature curves for uniform rates of change of temperature of one degree in five minutes and one degree per hour are shown as broken and dashed lines, respectively.

The solid lines indicate the results of a study of the steady-state volumes. The upper solid line connects the values of the steady-state volumes of the specimen after it has been cooled

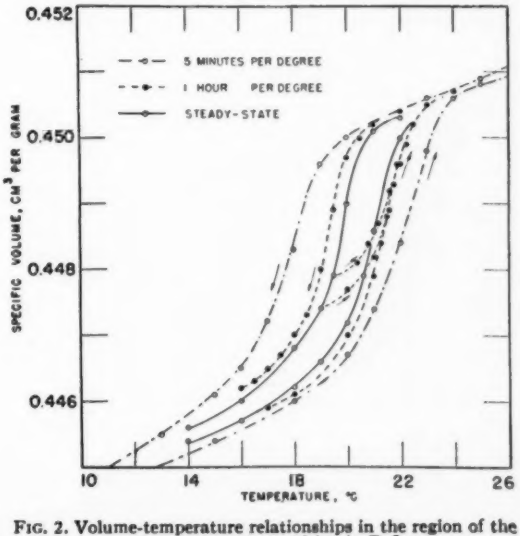


FIG. 2. Volume-temperature relationships in the region of the room temperature transition in Teflon.

rapidly through a 25°C interval to the temperatures indicated; similarly, the lower solid line connects the values of the steady-state volumes of the specimen after it has been heated through a 25°C interval. Thermal equilibrium as indicated by the attainment of the steady-state volume was reached in about 15 minutes when the 25°C interval did not include any part of the transition region. A period of about two hours was required for the attainment of the steady-state volume after heating or cooling through a 25° interval, which included all of the transition region. The longer time was required, in part at least, because of the latent heat of transition. However, intervals of time up to 60 hours in length were required to reach steady-state volumes upon cooling to temperatures located within the transition region. These intervals are far greater than could be needed for the attainment of thermal equilibrium. The length of time required depends on the location of the temperature within the transition region. On heating to temperatures within the transition region, steady-state volumes were attained in the length of time (never more than two hours) required for temperature equilibrium.

It should be emphasized that the dimensional changes resulting from the room temperature transition must be considered in the design of Teflon parts for uses involving close tolerances.

[1] H. A. Rigby and C. W. Bunn, Nature **164**, 583 (1949).
[2] C. E. Weir, National Bureau of Standards (unpublished).
[3] M. M. Renfrew and E. E. Lewis, Ind. Eng. Chem. **38**, 870 (1946).
[4] W. E. Hanford and R. M. Joyce, J. Am. Chem. Soc. **68**, 2082 (1946).
[5] N. Bekkedahl, J. Research Natl. Bur. Standards **43**, 145 (1949), RP2016.

Work-Hardening of Mild Steel by Explosive Attack

JOHN S. RINEHART
Michelson Laboratory, U. S. Naval Ordnance Test Station, Inyokern, China Lake, California
(Received May 11, 1951)

RAPID absorption of the high intensity compressional stress wave that is generated in a steel plate by application of an explosive charge has been established.[1] Appearance of at least part of this absorbed energy in the form of cold working of the metal is to be expected.[2] Since the initial stress is extremely high, about 300,000 atmospheres and of very short duration, two or three microseconds,[3] it is of considerable interest to observe the character of the work-hardening that takes place. The fact that the wave is attenuated as it passes through the steel permits observations to be made at varying stress levels.

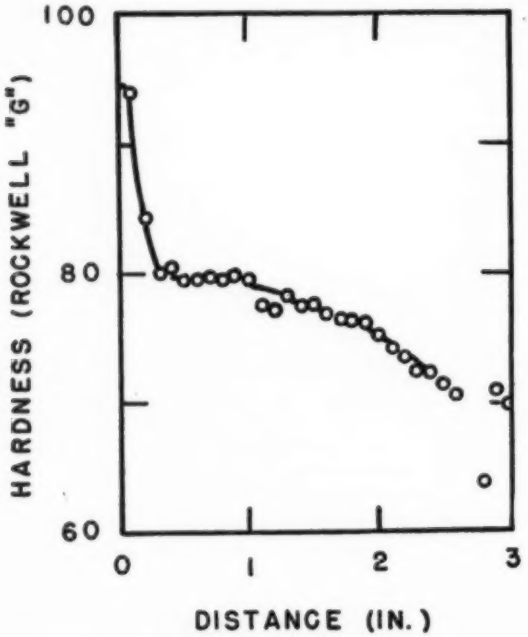


FIG. 1. Work-hardening curve for annealed 1020 steel plate. Thickness of layer of explosive, ½ in.

The hardness data reported here were obtained from medium 1020 steel plates, each of which had been acted upon by a thin layer of explosive that was detonated on the surface of the plate. Each plate was sectioned and the Rockwell hardness measured at several points across the section. Some of the specimen plates were cut from standard, unannealed, large, hot-rolled plates whose original hardnesses were not the same but differed by five or six points, Rockwell "*F*." Annealed plates were used in other tests. The thicknesses of the layers of explosives used ranged from ¼ in. to 2 in. The steel plates were 2 in. to 3 in. thick.

The data show that the change in hardness which has resulted from passage of the stress wave varies in a characteristic, but not simple, manner with the distance the wave has traveled. A typical curve that shows the general nature of this variation is plotted in Fig. 1. Initially, the hardness decreases rapidly with increasing distance from the explosive. This trend, however, stops abruptly; the curve turns sharply, and the hardness remains constant for a considerable interval. It then begins to decrease more or less gradually toward the original hardness of the plate.

The existence of a plateau in the curve was unexpected. Further, it was found that the difference between the hardness level of the plateau and that of the parent material was very nearly constant, about 10 points, Rockwell "*F*." Variations in the thickness of the explosive layer seemed to have no effect on the magnitude of the change in hardness. Increasing the thickness of the layer of explosive merely moved to the right the point at which the plateau began. For the ¼-in. layer of explosive, the plateau began at about 0.1 in., for the 2-in. layer, at about 0.5 in. The plateau was also somewhat longer for the thicker layers of explosive.

A number of general conclusions can be drawn as to how the change in hardness produced by the stress wave is influenced by the original or residual hardness of the parent material, although no systematic attempt has been made to investigate this phase of the problem. First, the point, with respect to distance, at which the plateau begins does not change significantly with original plate hardness. Second, the difference between the hardness level of the plateau and that of the parent material is constant, about 10 points, Rockwell "*F*," the same as was observed when the thickness of the explosive layer was changed. Third, high original plate hardness shortens the length of the plateau. And fourth, the depth in the plate to which a given stress wave will produce appreciable work-hardening increases with decreasing original plate hardness.

The abrupt change in slope of the hardness curve and the existence of a relatively flat plateau suggest strongly that two distinct mechanisms are operative in bringing about the observed changes in hardness. Such behavior would be entirely consistent with present ideas regarding the mechanics of the work-hardening of polycrystalline materials.[4] Upon being heavily etched, each section exhibited a thin layer adjacent to the face on which the explosive had been detonated that contrasted markedly with the remainder of the section. The lower boundary of this layer corresponded in every case with the beginning of the plateau of the hardness curve. The region below this layer was characterized by the presence of a large number of Neumann bands or shock twins.

The data and observations are also consistent in general with the view that more or less normal plastic flow, i.e., slip, grain breakdown, etc., superimposed on shock twinning accounts for the increased hardness in the narrow region close to the explosive and that shock twinning is mainly responsible for the increase in hardness throughout the rest of the plate. It is necessary to make certain assumptions in order to support this view and, in particular, to explain the occurrence and behavior of the plateau. First, there must be a limiting stress or, perhaps, impulse below which normal plastic flow will not take place. This assumption is in substantial agreement with Wood and Clark's observations[5] that the time required to initiate plastic flow in mild steel is a function of the applied stress. For the time of loading, a few microseconds,

involved here the stresses must be very high. Second, there must be an upper limit to the amount of shock twinning that can occur, and this limit will be reached at d stress or impulse considerably lower than that required to produce normal plastic flow. Geil and Carwile[6] have observed in their tensile tests of mild steel at low temperature that there appears to be a limiting amount of shock twinning that can occur. This observation is in substantial agreement with the first part of the present assumption. Third, high initial hardness of the plate tends to raise the stress or impulse required to produce shock twinning.

The occurrence of a plateau in the hardness curve seems to be peculiar to mild steel. Annealed copper, annealed brass, and 24S-T4 aluminum alloy have been investigated, but the hardness curves for these materials do not show abrupt changes in slope.

[1] J. S. Rinehart, J. Appl. Phys. (to be published).
[2] In this connection, it is interesting to note that Mr. Norman A. MacLeod has found that it is commercially feasible to work-harden Hadfield steel by explosive attack (private communication).
[3] See, for example, R. H. Kent, J. Appl. Phys. **13**, 348 (1942), and C. S. Robinson, *Explosions; Their Anatomy and Destructiveness* (McGraw-Hill Book Company, Inc., New York, 1944).
[4] A. M. Freudenthal, J. Franklin Inst. **248**, 6 (253) (1949).
[5] D. S. Wood and D. S. Clark "The influence of stress and temperature on the time for the initiation of plastic deformation in an annealed low carbon steel," Calif. Inst. Tech. Dynamic Properties Laboratory Report No. 24418-1, June, 1949.
[6] G. W. Geil and N. L. Carwile, J. Research Natl. Bur. Standards **45**, 129 (1950).

On Kerr Cell Construction

W. R. Smythe
California Institute of Technology, Pasadena, California
(Received April 30, 1951)

AN item in a Letter to the Editor[1] by J. W. Beams and H. S. Horton, together with a recent conversation with a user of Kerr cells, suggests that a very simple solution of the inhomogeneous field problem in such cells may have been overlooked. With a substance such as nitrobenzene whose specific inductive capacity is 36 a cell of small volume is needed to reduce the capacitance. With plates *inside* the cell this gives bad fringing electric fields. If, however, the cell is built of insulating material between two vertical flat parallel conducting plates, as shown in Fig. 1, with all dielectric boundaries either in contact with or normal to them, no distortion whatever of the uniform field will be produced by these boundaries. This is true because all equipotential surfaces parallel the plates and so give no displacement components normal to dielectric boundaries. In the two-dimensional case of two parallel semi-infinite thin oppositely charged conducting plates, the field has attained roughly 99 percent of its final value at a distance inside the edge equal to that between the plates. Thus for the cubical Kerr cell shown, if the plates extend beyond the cell walls for a distance equal to the length of a side, the variation in field strength in the liquid should be of the order of 1 percent. The flanges, which have 8 times the area of the cell but only 1/36 its capacitivity, add roughly 23 percent to its capacitance.

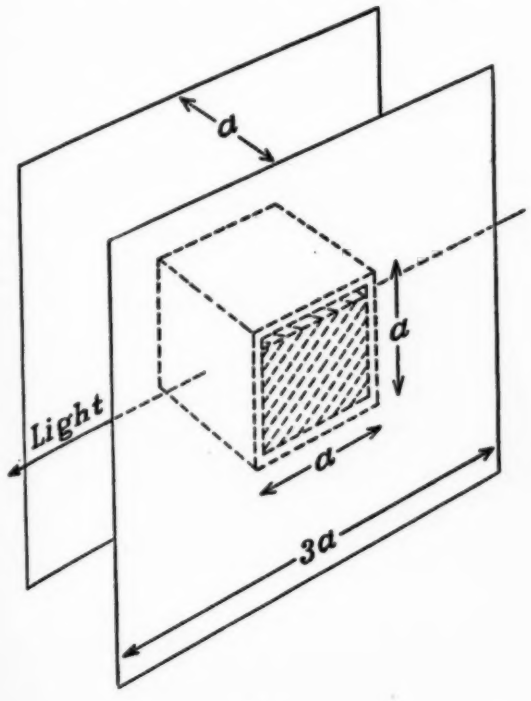


FIG. 1.

[1] J. W. Beams and H. S. Horton, J. Appl. Phys. **22**, 523 (1951).

A Study of Oxide-Coated Cathode by X-Ray Diffraction Method

Eiso Yamaka
The Electrical Communication Laboratory, Tokyo, Japan
(Received April 12, 1951)

NO direct measurements of the relation between crystal size and thermionic emission of BaO, (BaSr)O, and (BaSrCa)O have yet been carried out. It is desirable to make use of a vacuum tube in which both the crystal size and the emission can be measured simultaneously and directly. A method of cracking a tube and dipping a cathode into a protecting wax as done by Eisenstein[1] is not desirable because this method makes the repetition of measurements in the same cathode impossible.

Recently we have succeeded in making a vacuum tube with two mica windows and a movable anode, as shown in Fig. 1, and by using this tube the following measurements of the crystal size and the emission could be made. In these measurements the mica windows are used as entrance and exit ports for the x-ray beam. The intensity distribution curve of the (111) Debye-Sherrer line is measured by a G-M counter and scale-of-16. The crystal size is calculated from the half-line-breadth of the intensity curve by the well-known Sherrer equation. The emission is measured with pulses in order to minimize heating in the tube. An ionization gauge is attached to the test tube.

The measurements of pressure, emission, and crystal size are carried out in succession at several stages during heat treatment of the cathode, until both crystal size and emission reach the equilibrium values at a temperature (say T_1). When the equilibrium is obtained at T_1, the cathode is heated again at a higher temperature (T_2) and the measurements are repeated as before. Thus we obtain Fig. 2.

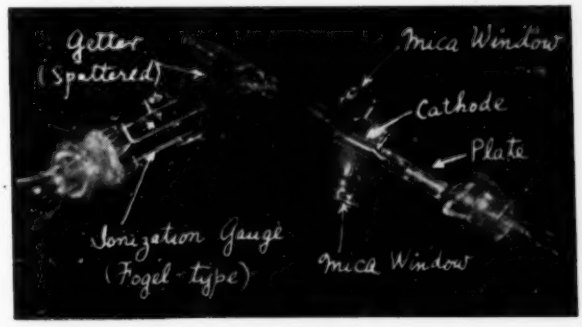


FIG. 1. Test tube.

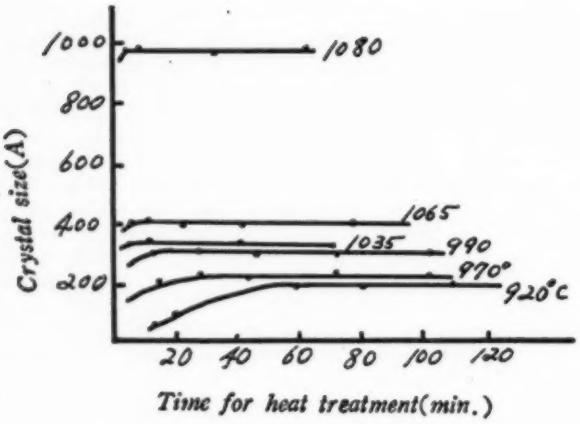


FIG. 2. Time of heat treatment *vs* particle size.

From the experiments carried out, it was found that a crystal size of (BaSr)O is about 200A at 920°C and grows slowly to 1050°C. At higher temperatures it grows very rapidly, as large as 1000A, as shown in Fig. 3. Also it can be mentioned that in the

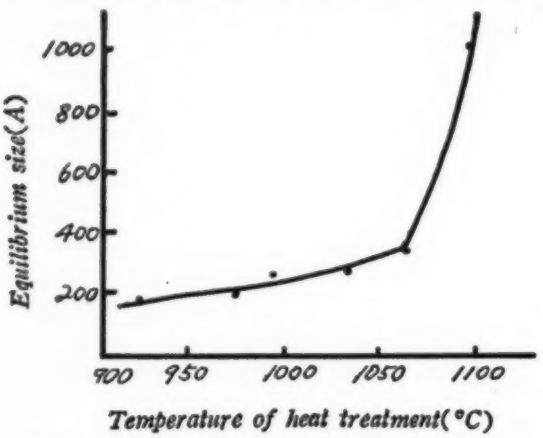


FIG. 3. Temperature of heat treatment *vs* particle size.

usual cathode containing impurities, the impurity activation is dominant at lower temperatures and the heat treatment dominant at higher temperatures. This arises because the curve of $\log N_0/T$ (curve A) is a nearly straight line in contrast to the convex downward curvature predicted from the theory[2] (curve B) of the relation between the crystal size and the emission, as seen from Fig. 4. N_0 is the density of thermionic active centers and is calculated from the Richardson curve.

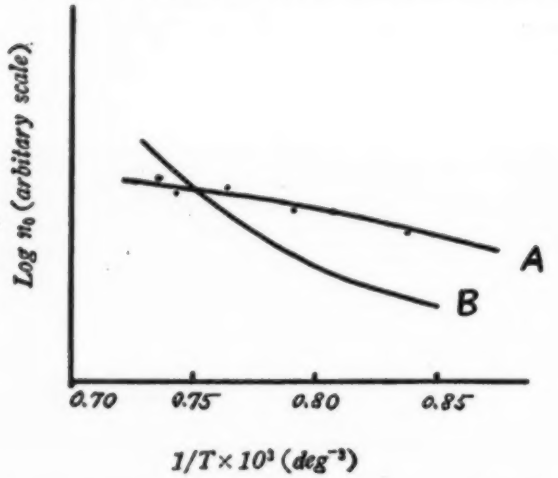


FIG. 4. $\log n_0 - 1/T$. (a) Experimental curve. (b) Theoretical curve.

Similar experiments on BaO and (BaSrCa)O are now in progress and the results will be reported later.

[1] A. Eisenstein, J. Appl. Phys. 17, 654 (1946).
[2] H. Kawamura and others, *Electron Emission and Semiconductors* (1949) (in Japanese), p. 245.

X-Ray Detection of Long-Range Order in Ni_3Mn

B. L. AVERBACH
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received May 18, 1951)

MAGNETIC and electrical resistivity data have indicated the possibility that an ordered structure can be produced in the alloy Ni_3Mn. However, it is difficult to detect long range order in this system, since the intensity of the superstructure lines is proportional to $(f_{Ni}-f_{Mn})^2$, where f_{Ni} and f_{Mn} are the atomic scattering factors, and this factor is quite small for nickel-manganese alloys.

A melt containing 25.5 atomic percent manganese (24.3 weight percent) was prepared from electrolytic manganese and Mond nickel. The ingot was homogenized at 1000°C, quenched and cold reduced 94 percent to a 1.5-mm wire. The specimen was ordered by heating to 600°C and cooling at 20°C per hour in the range 600–500°C and at 20°C per day in the range 500–300°C. Recrystallization also occurred as a result of this treatment and the grain size was fine enough to produce smooth x-ray diffraction lines. After the ordering treatment the diameter of the wire was reduced to 0.5 mm by electro-polishing.

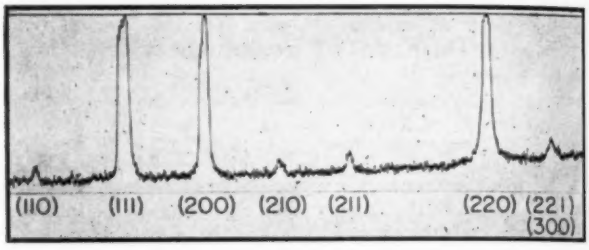


FIG. 1. Microphotometer trace of diffraction pattern from ordered Ni_3Mn. FeKα-radiation, fluorite monochromator.

X-ray diffraction photograms were made with FeKα-radiation in a Debye camera of 88.4 diameter. A fluorite crystal cut parallel to the (111) planes was used as a monochromator. Such a monochromator does not diffract the half-wavelength from the continuous spectrum, since the structure factor for the (222) planes is approximately zero. If this precaution is not taken, a reflection for each fundamental line is obtained from the half-wavelength at exactly the same Bragg angle as the superstructure reflection from the characteristic wavelength. For example, $\lambda/2$ would diffract the (200) and (220) fundamental lines at the same positions as λ would diffract the (100) and (110) superstructure lines. It thus becomes difficult to demonstrate the existence of a superlattice in this system where the half-wavelength reflections are about as strong as the superlattice lines. The one-third wavelength can also be diffracted, but this effect was minimized by operating the x-ray tube at 22 kv. The camera was evacuated to reduce air scattering.

Figure 1 is a microphotometer trace of a film produced by a 72-hour exposure under the above conditions. The superstructure lines (110), (210), (211) and the combination (300) and (221) are clearly visible. The (100) line could be seen, and its position determined on the film, but it was too weak to be recorded by the

microphotometer. The superstructure lines are consistent with a Cu_3Au ordered structure and a lattice parameter of approximately 3.59A referred to the disordered f.c.c. structure. The degree of order is now being measured as a function of temperature by a comparison of the relative intensities of the superstructure and fundamental lines.

The specimen used in this work was prepared and made available by L. R. Aronin of the M. I. T. Metallurgical Project in connection with an investigation of nickel-manganese alloys currently being performed under AEC Contract No. AT(30-1)-981.

A Remark on the Composition Dependence of the Elastic Parameters of Beta-Brass

JOHN S. RINEHART
Michelson Laboratory, U. S. Naval Ordnance Test Station, Inyokern, China Lake, California
(Received May 11, 1951)

RECENTLY Artman and Thompson have reported[1] a dependence of the elastic parameters of beta-brass on composition. The writer's previously determined values[2] are in general agreement with this observation; however, some uncertainty has existed as to the exact composition of the crystals, since only that of the stock material was determined. Two of the crystals used in the earlier investigation have been analyzed recently in order to remove the uncertainty. A standard wet method was used. The analysis is listed in Table I. A small loss of zinc during growth

TABLE I. Analysis of two beta-brass crystals used in author's investigation. Numbers are weight percents.

Element	Crystal 1	Crystal 2	Average	Stock material
Tin	Nil[a]	Nil	Nil	Nil
Lead	Nil	Nil	Nil	0.02
Iron	Nil	Nil	Nil	0.01
Aluminum	Nil	Nil	Nil	Nil
Copper	52.55	52.67	52.61	51.78
Zinc	47.23	47.15	47.19	48.19[b]
Undetermined	0.22	0.18	0.20	—

[a] Trace metals were not quantitatively determined owing to the small size of the sample.
[b] By difference.

of the crystals apparently occurred as predicted by Artman and Thompson.[1] The atomic percent copper in the crystals was found to be 53.41 percent, a value somewhat higher than the 52.47 percent reported previously[3] for the stock material. The increase in relative copper content is sufficient to explain the observed relation of the writer's results to the other sets of data.

[1] R. A. Artman and D. O. Thompson, J. Appl. Phys. **22**, 358 (1951).
[2] J. S. Rinehart, Phys. Rev. **58**, 365 (1940).
[3] Reference 2 lists the atomic percent of copper in stock material incorrectly as 52.8.

On Linear Expansion Coefficient and Melting Point of Metals

G. BONFIGLIOLI AND G. MONTALENTI
Istituto Elettrotecnico Nazionale, Torino, Italy
(Received April 9, 1951)

IN a recent letter published on the same subject,[1] besides other related questions, it is observed that for some elements of monometric system a relation may be assumed between linear expansion coefficient λ and melting point as follows:

$$\lambda(T_m)^{1/3}=\text{const}, \qquad (1)$$

where T_m is the melting point in °C.

Considering for the moment metals only, we can observe that, if we assume for T the absolute temperatures, we reach two important results, namely: (a) the experimental data may, with good approximation, be summarized by the equation,

$$\lambda T_m=\text{const}, \qquad (2)$$

which has the advantage of a precise physical meaning, i.e., the total expansion in percent at the melting point is constant for every crystalline network. (b) This constant is characteristic of crystal symmetry and grows with symmetry; i.e., metals with lower symmetry are the less stable ones for what concerns melting.

Table I illustrates these statements, which, as may be easily seen, have, no doubt, a satisfactory physical basis.[2,3] Analysis of Table I indicates that $\lambda T_m=2.43\pm0.44$ (mean deviation) for b.c.c.; 2.08 ± 0.22 for f.c.c.; 2.05 ± 0.10 for Hex. metals. The proximity of values of f.c.c. and Hex. c.p. lattices is, of course, obvious, because of the deep similarity of these two structures.

It is also to be observed that we have used expansion coefficients at temperatures very far from the melting point, because of the lack of more suitable experimental data; consequently, for metals with a high melting point defective values may be found for the constant λT_m characteristic of crystal symmetry; this because we have not considered the growth of λ with temperature (Feδ, Tiβ, Mo, Ir, Pt, etc.).

Taking into account such a limitation, we may verify that, even for high symmetries, mean values of λT_m are separated; and this fact, compared with the rapid decrease of the constant for the low symmetry classes (tetr., trig.) greatly enhances the opinion that the unavoidable scattering of experimental points masks a real feature of the phenomenon. This experimental result, that the crystalline network melts when it reaches a certain critical linear expansion of a few percent depending on symmetry and not on the atom, should be able, as we see, to throw some light on the causes of melting.

Finally, it seems interesting to observe that some more remarkable exceptions are encountered with metals undergoing allo-

TABLE I.

Metal	Symmetry	Melting point °K	λT_M in %	$\overline{\lambda T_M}$ For the class	Mean dev. of the class
Cs	b.c.c.	301.5	2.90		
Rb	b.c.c.	311.5	2.98		
K	b.c.c.	335	2.86		
Na	b.c.c.	370.5	2.75	2.43%	0.44
Li	b.c.c.	459	2.80		
Feδ	b.c.c.	1808	2.15		
Tiβ	b.c.c.	2073	1.89		
Mo	b.c.c.	2893	1.50		
Tlβ	f.c.c.	576.5	1.66		
Pb	f.c.c.	600.5	1.71		
Al	f.c.c.	933	2.06		
Ca	f.c.c.	1083	2.51		
Ag	f.c.c.	1233.5	2.32		
Au	f.c.c.	1334	1.90	2.08%	0.22
Cu	f.c.c.	1356	2.17		
Niβ	f.c.c.	1728	2.36		
Coβ	f.c.c.	1753	2.17		
Pd	f.c.c.	1826	2.08		
Pt	f.c.c.	2046.5	1.81		
Pr	f.c.c.	2623	1.70		
Cd	hex.	594	1.87		
Zn	hex.	693	2.10		
Mg	hex.	924	2.18	2.05%	0.10
Be	hex.	1623	2.16		
Os	hex.	2973	1.87		
In	tetr.	428	1.38	1.30%	0.05
Snβ	tetr.	505	1.22		
Bi	trig.	544.5	0.75		
Sb	trig.	903.5	1.18	0.93%	0.12
As	trig.	1087	0.87		

tropic changes (Mo, Fe, Tl, etc.). It consequently does not seem impossible that some "scattered" points relative to metals until the present date considered as nonpolymorphic, may hide unknown allotropic changes. In this sense the experimental value of λT_m could be considered as an index of possible transformations. Some further remarks should be made:

(1) Data chosen for λ (and probably also values chosen by M. E. Straumanis) concern polycrystalline materials. On the contrary every side of the elementary cell has its own total expansion at melting point, and it could perhaps be more significant to consider, for every metal, the smallest of the different expansions. It should be of some interest, if the data available on monocrystals were not so few, to verify the effect of this way of plotting results. Moreover, it is probable that differences will not be very high and such to change radically the above statements.

(2) The following remarks concern the scattering of various values of Table I.

W—It has not been put therein because, at 950 °K, it transforms to $W\beta$ cubic complex. As a matter of fact, it is believed that the normal form is the stable one at melting, if special conditions are not created. Attention besides is drawn to the fact that the value of λT_m constant is in accordance with a lower symmetry.

Cr—A complex cubic structure is probable at high temperature.[4] The experimental value confirms this hypothesis.

Ca—An allotropic transfer above 650°K is probable.[5]

Mn—Not marked for the complex allotropic changes jointly with great variations of λ.[6]

Zr—Not marked because of the uncertainty of $\pm 700/2500$°K on T_m.[7]

Ru, Cb, Ga, Se, Si, V—Not marked because λ was uncertain or unknown.

Fe—Choosing data from reference 4, p. 427, for $\lambda(T)$ and integrating to T_m we find $\lambda T = 2.7$ percent; this proves that some exceptions to melting at elevated temperature could probably be corrected in this way.

This kind of reasoning, at first sight unjustified if not applied to every point of Fig. 1 of reference 4, is, as a matter of fact, not so arbitrary, because $d\lambda/dT$ has the same order of magnitude for all pure metals, and consequently our correction is expected to be significant only for metals melting at high temperature.

Ir, Pt, Rh—An allotropic transformation is probable for Rh.[8] It is not impossible for the other two elements.

Pb, Bi—Perhaps low melting temperature and high plasticity would give some trouble on determination of λ and evaluation of λT_m constant.

In the above remarks, all pure metals have been considered.

[1] M. E. Straumanis, J. Appl. Phys. 21, 936 (1950).
[2] Data for Table I have been taken chiefly from: Mott and Jones, *Theory of the Properties of Metals and Alloys* (Cambridge University Press, London, 1936); and completed, when necessary, with data taken from the *Handbook of Metals* (1948).
[3] Elements showing structure variations have been enumerated on the basis of their form stable at high temperature.
[4] *Handbook of Metals* (1948), p. 1135.
[5] See reference 4, p. 1134.
[6] See reference 4, p. 1138.
[7] See reference 4, p. 20.
[8] See reference 4, p. 195.

Announcements

Nuclear Physics in Europe

Six of the distinguished European physicists who are coming to this country in September will conduct a two-day symposium at Oak Ridge, Tennessee, on September 13 and 14 on the general subject "Nuclear Physics in Europe."

Symposium leaders and their subjects are as follows:

Professor E. Amaldi, University of Rome, "Contribution to the Investigation of Extensive Showers."

Professor J. Rotblatt, University of London, "Studies of Nuclear Processes Involving 8-Mev Deuterons by the Photographic Method."

Professor R. E. Peierls, University of Birmingham, "Problems of Nuclear Forces and Nuclear Structure."

Professor S. Devons, Imperial College of Science and Technology, London, "Projected Experiments on Gamma-Emission."

Professor P. Huber, University of Basel, "Elastic Scattering of Fast Neutrons from Various Light Elements."

Professor J. Mattauch, University of Bern, "Problems of Modern Mass Spectrometry."

The Oak Ridge symposium is unrestricted in nature, and all interested physicists and others are invited to attend.

Additional information may be obtained from the University Relations Division, Oak Ridge Institute of Nuclear Studies, P. O. Box 117, Oak Ridge, Tennessee.

1951 Awards for Essays on Gravity

INSTEAD of seven awards, as in 1950, we are now offering eight awards for short essays with the following six stipulations.

(1) These eight awards will be made by us on November 23, 1951, for the best 1500-word essays* (a) on the possibilities of discovering some partial insulator, reflector, or absorber of gravity, or (b) on the possibilities of discovering some alloy, the atoms of which can be agitated or rearranged by gravity tension to throw off heat, or (c) on the possibilities of discovering some alloy the temperature of which can be affected by gravity waves.

(2) The first award will be $1000.00. The second award will be $400.00. The third award will be $200.00. The fourth award will be $150.00. There will also be four $100.00 awards.

(3) Essays must be received at this office before October 15, 1951. They will be accepted from anyone who is seriously interested in the application of gravity to practical uses for the benefit of humanity.

(4) All essays must be typewritten on paper 11×8½ inches, with two carbon copies.

(5) Those who have submitted essays in previous years, but received neither First nor Second Awards, are free to compete again.

(6) The decision of the Trustees will be final, and all essays and copies will become the property of the Foundation.

Address:
GRAVITY RESEARCH FOUNDATION,
NEW BOSTON, NEW HAMPSHIRE
GEORGE M. RIDEOUT, PRESIDENT

* It may be less than 1500 words; but if more the judges have the right not to accept same.